THE PHYSICAL SOCIETY

Reports on

PROGRESS IN PHYSICS

VOLUME VI

GENERAL EDITOR
J. H. AWBERY

Price to non-Fellows 22s. 6d. net

Reprinted by
WM. DAWSON & SONS LTD.,
with the permission of the Publisher

Reprinted in The Netherlands

CONTENTS

PAGE

The physics of stellar interiors and stellar evolution. By H. A. BETHE and R. E. MARSHAK 1

Induced radioactivity. By [The late] H. J. WALKE 16

Separation of isotopes. By H. C. UREY 48

The meson. By R. PEIERLS 78

The production and measurement of short-wave radiations. By G. W. C. KAYE, G. E. BELL, W. BINKS and W. E. PERRY . . 95

The cyclotron and some of its applications. By W. B. MANN . . 125

Spectroscopy. *Editor*, W. JEVONS 137

I. The spectra of supernovae. By A. HUNTER

II. Some recent developments in the Zeeman effect in atomic spectra. By H. E. WHITE

III. The magnetic effect in diatomic spectra. By F. H. CRAWFORD

The luminescence of solids. By J. T. RANDALL and M. H. F. WILKINS 174

Reactions in solids. By N. F. MOTT 186

The theory of molecular structure By W. G. PENNEY 212

Sound. By E. G. RICHARDSON 238

Fluid motion. Transition from laminar to turbulent flow in a boundary layer. By A. FAGE 270

The properties of liquid helium. By H. JONES 280

Heat. By R. W. POWELL 297

Kinetic theory of the elasticity of rubber. By H. PELZER . . . 330

Superconductivity. By L. C. JACKSON 335

PAGE

X rays and crystals. *Editor*, J. M. ROBERTSON 345

I. Organic structures. By J. M. ROBERTSON

II. The structures of silicates. By W. H. TAYLOR

III. The examination of ternary alloy systems by x rays. By H. LIPSON

IV. Order-disorder phenomena in alloys. By E. T. GOODWIN

Measurements with ultra-short electric waves on wires. By L. HARTSHORN 378

Impedance networks. By R. S. RIVLIN 389

The measurement of capacitance. By N. F. ASTBURY 403

Dielectric breakdown in solids. By H. FRÖHLICH 411

The teaching of physics in technical institutions. By D. OWEN . . 431

THE PHYSICS OF STELLAR INTERIORS AND STELLAR EVOLUTION

By H. A. BETHE and R. E. MARSHAK *
Cornell University, Ithaca, N.Y.

During the last two years, numerous theories have been advanced in the attempt to explain the mechanism of energy production in various types of stars. These theories, based as they are on our vastly increased knowledge of nuclear phenomena, have been developed not only qualitatively but to a considerable extent quantitatively. In this manner it seems to have been well established that thermo-nuclear processes are eminently capable of explaining the energy generation, stability and empirical regularities of "normal" main-sequence stars and possibly even the giants, but cannot account for the "abnormal" behaviour of such stars as novae, supernovae, and white dwarfs. The energy production in white dwarfs is most probably due to gravitational contraction, while the suggestions made thus far to explain novae and supernovae are purely speculative. In addition, these investigations have had some interesting consequences on our belief as to the origins of stars and their future courses of evolution. In the following pages we shall attempt to summarize briefly the general picture which we now possess concerning these matters.

§ 1. NATURE OF THE OBSERVATIONAL MATERIAL

Before we proceed to give an account of the recent theoretical developments, it seems worth while to review briefly the type of astronomical data which serve as the starting point for the theory, and to note the empirical correlations which are revealed and must be explained by any successful theory. The fundamental quantities which characterize the configurations of matter known as stars are the total mass M, total luminosity L, and total radius R. For a direct determination of the masses of individual stars it is necessary to have recourse to binary stars. The total mass of a visual binary system may be obtained by one of several methods, e.g. by the direct parallax method used in conjunction with a knowledge of the relative orbits; proper-motion data permit an assignment of the separate masses. For spectroscopic binaries which are also eclipsing variables, a separate assignment of the masses may also be made by a study of the velocity curves and light variations. The luminosity itself is never directly known from observation but usually from the absolute visual (or photo-visual) magnitude taken together with a table of bolometric corrections. The radii can be determined directly for double-line spectroscopic binaries which are also eclipsing variables and by means of interferometric measurements for some of the giants; but, in general, the radii

* Now at the University of Rochester, Rochester, N.Y.

are derived from a knowledge of the effective temperature on the surface of the star, T_{eff} (itself related to the spectral class of the star as revealed by its spectrum) and by use of the formula: $L=4\pi R^2.\sigma T_{\text{eff}}^4$. For more detailed information concerning these matters, it is suggested that the reader consult a paper by Kuiper [1].

When the data for stars for which M, L, and R are known are plotted on a mass-luminosity diagram and a mass-radius diagram, certain regularities are immediately manifest: most of the observed stars belong to the so-called *main sequence*, their luminosities and radii increasing fairly regularly with the mass. Because the luminosity increases at a more rapid rate with mass than does the radius, it can easily be seen that the effective temperatures will increase with mass, so that stars of this group will range from the hot and luminous blue giants to the cool and faint red dwarfs. It is to be supposed that these regularities are explicable in terms of a uniformity in the internal constitution of these stars and in the physical processes of energy production. At the upper end of the main sequence there exists a group of stars, the so-called Trumpler stars, which are very luminous O and B stars in galactic clusters and whose masses (determined by the Einstein gravitational red shift) are of the order of one hundred times that of the sun; these stars possess luminosities which depend much less strongly on the mass than do those of the less massive stars of the main sequence.

In addition, it is found that some stars of comparatively large mass (above three times the mass of the sun) exist in an entirely different state—one in which their luminosities are about the same as (or probably somewhat smaller than), and the radii much larger than, stars of the same mass in the main sequence. In such cases the effective temperatures are considerably lower and the stars are known as *red giants*. In the R–L diagram these stars form a sequence of their own (giant branch) almost perpendicular to, and branching off from, the middle of the main sequence. The Cepheid and Mira variables, stars which exhibit periodic luminosity changes (ascribed to mechanical pulsations of the entire star) represent the upper boundary of the giant branch with respect to L in the R–L diagram.

On the other hand, there exists a group of stars known as the *white dwarfs* with masses equal to that of the sun or smaller, but which possess much smaller luminosities and radii than the corresponding stars of the main sequence. Probably closely related to the white dwarfs are the central stars of the planetary nebulae (Wolf-Rayet stars), which have smaller radii for given luminosities.

Finally, astronomers have discovered objects in the sky—novae and supernovae—which liberate tremendous amounts of energy in extremely short periods of time *; these outbursts are at present unexplained (cf., however, reference [22]).

§ 2. THE THEORETICAL PROBLEM OF STELLAR CONSTITUTION

According to an old hypothesis of Vogt [2] and Russell [3], the constitution of a star in equilibrium is completely determined by its mass and chemical

* The behaviour of supernovae is described in some detail in the "Spectroscopy" *Report*, p. 137. [EDITOR.]

composition. From the observational material it seems that, within certain limits, even the chemical composition is of minor importance, since the radii and luminosities of main-sequence stars depend essentially only on their masses. On the other hand, the existence of the red giants and the white dwarfs shows that the mass alone does not determine the stellar constitution, and it seems most plausible to assume that these stars differ from the main sequence in chemical composition. This hypothesis is in accord with modern theories (§ 4 and § 5).

A complete theory of stellar constitution must fulfil the programme set by the Vogt-Russell hypothesis, i.e. it must permit a calculation of all stellar data from the mass and chemical composition. The most important data directly amenable to observation are, of course, the radius and luminosity. Besides, the distributions of density and temperature throughout the star are of theoretical interest.

As a basis of such a theory, information about three points is required:

(1) The behaviour of matter at high pressures and temperatures;

(2) the mechanism of transport of energy through the star;

(3) the mechanism of energy production.

The first two points have been virtually solved by astrophysicists during the last fifteen years [4], [5]. An essential help in this solution was the calculation of important atomic constants from wave mechanics (degree of ionization, opacity, etc.). The third point, on the other hand, could only be solved in 1938 by applying our greatly increased knowledge about nuclear physics, especially about nuclear transmutations.

Until 1937, the energy production remained the great unknown. It was therefore only possible to calculate *one* of the important observables of a star (e.g., its luminosity) when the other (the radius) was given (in addition to mass and chemical composition). Even for this programme it was necessary to know the *distribution* of the sources of energy through the star in order to deduce the flux of energy (point (2) above) at each point. This led to the investigation of stellar "models", of which the most important were Eddington's "standard model" and the "point-source model". In the standard model, the sources of energy were assumed to be distributed almost uniformly through the star; in the point-source model practically all the energy was supposed to be produced near the centre. As we know now, the second assumption is very near the truth.

Fortunately, it turned out that the characteristics of the star did not depend sensitively on the model. Moreover, the luminosity was shown to be almost independent of the stellar radius and to depend only on mass and chemical composition. At first the latter dependence was not realized, and it was assumed that stars contained mainly heavy elements (heavier than helium), with the result that the theoretical luminosity came out up to several hundred times larger than the observed. Eddington showed that this discrepancy could be removed by assuming a large content of hydrogen and/or helium, thus reducing the mean molecular weight of stellar matter which enters

the theoretical expression for the luminosity to a high power (about the seventh). Strömgren was then able to calculate the hydrogen contents of many stars from the observed luminosities, assuming the helium content to be small.

Such calculations yield, among other data, the central temperatures of the stars, which turn out remarkably constant in the main sequence. They vary from about 20 million degrees for the sun to slightly over 30 for the blue giants, while the luminosities change by a factor of about 10,000. This result shows that the energy production must be extremely sensitive to temperature, as indeed is actually the case for the nuclear reactions now believed to be responsible for the energy production.

For the red giants, similar calculations gave much lower central temperatures, ranging from 5 million degrees downwards. This shows again that the mechanism of energy production must be different in these stars.

With the solution of the problem of energy production by nuclear reactions it is now possible to say theoretically that the central temperatures of the majority of stars must have values of the same order of magnitude as those found by astrophysicists for main-sequence stars, viz. 20 to 30 million degrees, and must change with the mass of the star just in the way found in the astrophysical calculations. This means essentially that we are now able to predict correctly by theory alone the radius as well as the luminosity of a main-sequence star, and that we have thus proved the Russell-Vogt hypothesis to be correct. Even the lower central temperatures of the giants seem to have a definite significance in terms of nuclear reactions (§ 5).

§ 3. DESCRIPTION OF THE QUANTITATIVE THEORY

In the quantitative theory, the three points mentioned in the preceding section must be discussed further.

(1) The distribution of matter is governed by the equations of *hydrostatic equilibrium* [4], viz.

$$dp/dr = -(GM_r/r^2)\rho, \qquad \ldots.(1)$$

$$dM_r/dr = 4\pi r^2\rho, \qquad \ldots.(2)$$

$$M_r = 0 \text{ for } r = 0,\ M_r = M \text{ for } r = R,$$

where

r = distance from centre of star of radius R,
ρ = density,
p = pressure,
G = gravitational constant,
M_r = mass inside r.

The pressure is the sum of gas pressure and radiation pressure,

$$p = p_G + p_R, \qquad \ldots.(3)$$

of which the latter is given by

$$p_R = \tfrac{1}{3}aT^4 \qquad \ldots.(4)$$

(a = Stefan-Boltzmann constant), and is negligible for all but the heaviest stars. The gas pressure is a function of density, temperature and chemical composition, as follows:

(a) In all the brighter stars, i.e. all stars except the white and possibly some

of the red dwarfs, the gas is *non-degenerate* and therefore follows the usual gas law,

$$p = k\rho T/m_H\mu, \qquad \ldots.(5)$$

where

T = temperature (absolute),
k = Boltzmann's constant,
m_H = mass of hydrogen atom,
μ = mean molecular weight.

To obtain μ, the degree of ionization must be known. This can easily be calculated using quantum statistics and the known ionization potentials of the atoms concerned. It is found that in the interior of all ordinary stars the ionization is practically complete except for the K-shell of heavy elements such as Fe. Then, for an atom of atomic weight A having Z electrons,

$$\mu = A/(Z+1), \qquad \ldots.(6)$$

so that $\mu = 1/2$ for H, $\mu = 4/3$ for He, and $\mu \simeq 2$ for all the heavier elements. For a mixture of elements the average of $1/\mu$ must be taken.

If (5) is valid, the central temperature T_c of the star is approximately given by

$$kT_c \simeq GM\mu/R, \qquad \ldots.(7)$$

i.e. the average kinetic energy is equal to the potential energy of an average molecule in the gravitational field of the star.

μ is the quantity through which the chemical composition of the star enters most strongly into the calculations of luminosity and other stellar quantities. From the observed values of M, R and L the value of μ can be determined with good accuracy. For the sun $\mu \sim 1$, which corresponds to a hydrogen content of 35 per cent, assuming no helium. For most of the blue giants the hydrogen content is much higher ($\sim$80 per cent).

(*b*) In very dense stars (white and centre of some red or sub-dwarfs [5a]) the electron gas becomes degenerate. If the degeneracy is strong (very high density), the gas pressure depends only on the electron density ρ_e, which can easily be calculated from the mass density ρ in view of the fact that there is one electron per mass m_H in hydrogen, and per mass $2m_H$ in all other elements. p_G is in this case practically independent of the temperature.

(2) The energy transport may be by radiation, by electronic heat conduction or by convection.

(*a*) In all ordinary stars (main sequence and red giants) the *energy transport is by radiation* over the larger part of the star. Then the temperature distribution is determined by the equations of radiative equilibrium [4],

$$dp_R/dr = -K(\rho, \mu, T).\rho/c.L_r/4\pi r^2, \qquad \ldots.(8)$$

$$dL_r/dr = 4\pi\rho r^2\epsilon, \qquad \ldots.(9)$$

$$L_r = 0 \text{ for } r = 0,\ L_r = L \text{ for } r = R,$$

where K = opacity,
L_r = flux of energy through sphere of radius r,
ϵ = energy production per gram per second.

The opacity K of stellar matter for radiation is a function of density, temperature and chemical composition. At moderate temperatures and densities the opacity is mainly due to absorption of light by bound electrons, at higher temperatures or very high densities absorption by free electrons becomes more important, and at extremely high temperatures Compton scattering by free electrons is dominant. Whatever the source of the opacity, it can be calculated with considerable accuracy from quantum mechanics. Tables of K as a function of ρ and T were computed by Strömgren [5].

The energy production ϵ will be discussed below. Before the theory of ϵ was known, arbitrary assumptions had to be made about ϵ as a function of r, given the total energy production L. In the standard model it was assumed that KL_r/M_r was constant throughout the star; in the point-source model, $L_r=L$ for all values of r outside the convective core (see sub-section (*c*)).

(*b*) *Electron conduction* is most important for stars of very high density (white dwarfs), as was first realized by Kothari [6]. It can also be calculated from quantum theory, and equations (8) and (9) remain unchanged in form.

(*c*) Near the centre of normal stars, energy is transported by *convection*. This will always occur when the temperature gradient obtained from (8) and (9) becomes greater than the adiabatic gradient. (For details, see Strömgren [5].) If the gas is monatomic and the radiation pressure negligible, (8) is replaced by

$$d \log P/d \log \rho = 5/3. \qquad \ldots . (10)$$

Otherwise, this equation is replaced by a more complicated one.

(3) The energy production in main-sequence stars and red giants is now believed to be due to reactions between the swiftly moving atomic nuclei in the hot gas in the star (thermo-nuclear reactions). The calculation of ϵ for thermo-nuclear reactions was originally performed by Atkinson and Houtermanns [7], and an improved formula has been given recently by Gamow and Teller [8]. The procedure is simply the following: the effective cross-section for nuclear reactions involving charged particles is given essentially by the product of: (1) the collision cross-section; using plane waves for the particles, this is $\Lambda^2/4\pi$ for central collisions, where Λ is the de Broglie wave-length of the incident particle with respect to the centre of mass of the combined system, and goes down rapidly for higher orbital momenta l as $(r_0/\Lambda)^{2l}$, r_0 being the nuclear radius; (2) the transparency of the Coulomb barrier for the incident particle given by

$$\exp\left\{-\frac{2\sqrt{m}}{\hbar}\int_{r_0}^{Z_1Z_2e^2/E}\left(\frac{Z_1Z_2e^2}{r}-E\right)^{1/2}dr\right\}.$$

In this expression m is the reduced mass of the system, E the relative energy, Z_1e, Z_2e the charges of the two reacting nuclei; and (3) the "width" for the particular nuclear reaction considered, i.e. essentially its probability per unit time when the reacting nuclei are together. For particle reactions this width can be derived from the observed cross-sections and for γ-ray emission from the observed resonance captures or approximately from the dipole or quadripole formulae [9]. The effective cross-section is now a function of the relative energy

of the two interacting nuclei, and must be integrated over their thermal distribution of velocities. When this is done one gets a formula for the total number of processes per unit mass per unit time for each reaction, which when multiplied by the energy evolution per process gives the rate of energy generation corresponding to a specified temperature. Thus in general it is found that

$$\epsilon(Z_1 Z_2) = \text{constant} \cdot X_1 X_2 \rho T^{-2/3} \cdot \exp\{-\text{constant}\,(Z_1^2 Z_2^2/T)^{1/3}\}, \quad \ldots . (11)$$

where X_1, X_2 are the concentrations of the two interacting nuclei. From the dependence of ϵ on the charge it is readily seen that the probability of the reaction decreases very rapidly with charge, so that protons interacting with light nuclei will play the most important part in the production of energy. For the capture of neutrons the above derivation does not apply, but free neutrons cannot be present in a star under ordinary conditions of temperature and density, so that the above expression for ϵ suffices for the purpose of our discussion.

§ 4. ENERGY PRODUCTION IN MAIN-SEQUENCE STARS

As has been pointed out at the end of the last section, the nuclear reactions most important for the energy production will be those involving protons. These reactions may conveniently be divided into two classes, viz.

(*a*) Reactions in which only protons are consumed.

(*b*) Reactions in which other nuclei also are consumed.

According to the astrophysical evidence, protons are much more abundant than any other nucleus in stars. Therefore reactions of class (*b*) can supply energy only for a very limited time, especially if the other nuclei consumed are of such rare elements as Li, Be, B—and only these light nuclei will react fast enough with protons to give any appreciable energy production at all. It seems likely that reactions of this type are responsible for the energy production in red giants (§ 5). For the most frequent type of stars, those of the main sequence, we must look for a more lasting source of energy, which can only be found in reactions of class (*a*).

It can easily be shown that the combination of protons will not lead to the building up of elements heavier than helium because if any heavier elements are formed they will be immediately destroyed by further reaction with protons which will lead back to helium (§ 5 and Bethe [9]). The combination of protons into α-particles can go essentially only in two ways, viz. (A) the direct combination of two protons and (B) with the use of carbon nuclei as catalysts (carbon cycle).

(A) The first mechanism starts with the interaction of two protons, which leads to the formation of a deuteron and the emission of a positron, viz. $H+H=D+e^+$. The probability of this process has been calculated [10], [11] using the Fermi β-decay theory, the Gamow-Teller selection rules and the average of a set of empirically determined decay constants. The deuteron is then transformed into helium by a series of fast processes involving further capture of protons. The reaction $H+H=D+e^+$ is intrinsically improbable since it involves the emission of a β ray, a process ordinarily about 10^{15} to 10^{25} times

slower than other nuclear reactions. But the reaction is favoured by the extremely low potential barrier between the reactants, and therefore becomes important in cooler stars (cf. equation (11)). Even in the sun it may contribute as much as one-half of the total luminosity, while in red dwarfs it becomes predominant. Observational data on red dwarfs [12] are meagre, but they agree, at least to the order of magnitude, with the theoretical prediction from the proton-proton reaction, as can be seen from the following examples :

Star	$M/M_\odot$	$\epsilon_{av.}$ (theor.)	$\epsilon_{av.}$ (observ.)
Castor C_2	0·57	5	0·2
μHer. BC	0·45	0·3	0·2
$-8°$ 4352 AB	0·35	0·1	0·2

M is the mass of the red dwarf, $M_\odot$ is the sun's mass, and $\epsilon_{av.}$ is the energy production averaged over the mass of the star, the " observed " value being simply L/M. The agreement seems significant in view of the fact that the carbon cycle would give only 10^{-4} ergs/g.sec. for the lightest star mentioned. However, the observational data are too scarce and the theory too incomplete (partial ionization and incipient degeneracy have not been considered) to draw any conclusions about details such as the dependence of the luminosity on the mass of the star.

(B) The second mechanism, which serves as a source of energy by effectively transforming into radiation some of the binding energy of the α-particle with respect to four protons and two electrons, uses carbon and nitrogen in the catalytic chain of reactions :

$$\begin{aligned} &{}^{12}\mathrm{C}+\mathrm{H}={}^{13}\mathrm{N}+\gamma\,; \quad {}^{13}\mathrm{N}={}^{13}\mathrm{C}+e^+, \\ &{}^{13}\mathrm{C}+\mathrm{H}={}^{14}\mathrm{N}+\gamma, \\ &{}^{14}\mathrm{N}+\mathrm{H}={}^{15}\mathrm{O}+\gamma\,; \quad {}^{15}\mathrm{O}={}^{15}\mathrm{N}+e^+, \\ &{}^{15}\mathrm{N}+\mathrm{H}={}^{12}\mathrm{C}+{}^{4}\mathrm{He}. \end{aligned}$$

The energy production due to this set of nuclear reactions is given [9] by a formula of type (1) with X_1, Z_1 referring to hydrogen and X_2, Z_2 to the carbon-nitrogen group. The catalyst ^{12}C is reproduced in all cases except about one in a hundred thousand, when the last reaction leads to ^{16}O instead. Therefore the abundance of carbon and nitrogen remains practically unchanged (in comparison with the consumption of protons).

If we assume for carbon a concentration of about 1/2 per cent by weight, then we can calculate the theoretical energy production and compare it with the observed for stars more massive than the sun. The following table gives a fair representation of the results :

Star	M/M	$\epsilon_{av.}$ (theor.)	$\epsilon_{av.}$ (observ.)
TX Her. f.	1·8	4	12
Sirius A	2·5	150	28
U Ophui br.	5·4	25	180
Z Vul. br.	7·6	250	330
Y Cyg. br.	17·3	1500	1200

It is seen that the agreement is quite excellent ; particularly is this seen to be the case when calculations are made using two quite different models, namely,

the standard and point-convective models (cf. § 2 and § 3). The point-convective model, which gives a much larger concentration of energy sources near the centre of the star than does the Eddington standard model, leads to a central temperature which is about 5 per cent greater than that given by the standard model; but it leads at the same time to a lower central density and to a smaller ratio of average to central energy production, which just about compensates the effect of the higher temperature. This implies that the results given in the table are accurate (say at most to a factor 2) as far as the stellar model is concerned.

About the only uncertainties which remain for these stars are the helium content and the concentration of reactants in the carbon cycle. The helium content influences the opacity, and therefore the temperature, to a considerable extent, so that it should be taken into consideration. As regards the presence of C and N, the observational evidence is very indeterminate because atmospheric conditions are such that their lines are diffioult to detect. An estimate of about a half per cent each for carbon and nitrogen seems quite compatible with spectroscopic observations. For carbon this is the same abundance as was assumed by Bethe [9]; for nitrogen a much higher abundance (10 per cent) was then assumed, a change which must be compensated by assuming for the intrinsic probability (width) of the reaction $^{14}N+H=^{15}N+\gamma$ a value which is twenty times larger and still quite plausible. The relative abundance of ^{14}N and ^{15}N also speaks in favour of the new assumptions about abundance and width.

These uncertainties will appear insignificant when the carbon cycle is compared with other nuclear reactions. At a temperature of 20 million degrees, such as is obtained from the astrophysical calculations in the preceding section, the next lighter element, boron, would give an energy production at least 10,000 times too large, the next heavier, oxygen, about 100,000 times too small. This automatically rules out oxygen, and *à fortiori* all heavier elements, as contributors of any importance to the energy production, while it shows that all elements between helium and carbon must have ceased to exist to any appreciable extent in main-sequence stars. That this is so is of course due to the fact that they are consumed when reacting with protons (class (*b*) reaction, cf. above and § 5), so that they disappear in the early history of the star. This conclusion does not apply to ^4He because ^5Li does not exist; however, the very non-existence of ^5Li prevents the interaction of helium and protons from contributing to the energy production. The carbon cycle is unique in that it is truly cyclical in character, leading to a reproduction of the original nucleus, and in that it gives the right energy evolution at temperatures which are consistent with the equations of stellar equilibrium.

The most conclusive evidence for the carbon cycle is perhaps that astrophysical data demand a very strong temperature dependence for the brighter (and therefore more massive) stars of the main sequence. This is just what the carbon cycle gives—it predicts a variation of ϵ about as T^{17}—while the only other possible source, the proton-proton reaction, does not give a strong enough dependence ($\epsilon \sim T^4$). Both Gamow [13] and Russell [14] have shown by statistical

considerations—using homology transformations in conjunction with the *R–M* and *L–M* diagrams of a well-known group of main-sequence stars—that a dependence of the energy on a high power of the temperature (about 16) gives splendid agreement with the observations.

It can also be shown that the strong temperature dependence of the reaction rate leads to a slight dependence of the central temperature on the mass and to the correct increase of radius and luminosity with mass and with molecular weight—in agreement with the integration of the star equations and with observation. Moreover, it can be shown that stars up to at least twenty times the mass of the sun have sufficiently low radiation pressure for them to be vibrationally stable in spite of the strong temperature dependence of the energy production [15].

We may therefore say that it follows conclusively from present knowledge (in spite of some minor uncertainties in the observational data) that the carbon cycle is responsible for the energy production of the main-sequence stars brighter than the sun. With somewhat less assurance we may state that the proton-proton reaction is the source of energy for main-sequence stars fainter than the sun. For the sun itself, theory predicts that the carbon cycle and the proton-proton reaction should give about equal contributions [11].

§ 5. ENERGY PRODUCTION IN THE RED GIANTS

In the previous section we have seen that the energy production of the main-sequence stars can be very well explained by means of the proton-proton reaction and the carbon cycle. However, for the red giants, for which any reasonable stellar model gives central temperatures in the neighbourhood of five million degrees (instead of the twenty million degrees for the main-sequence stars) and considerably lower densities, neither of these reactions can account for the observed high luminosities.

Gamow [16] has also shown that it is impossible to assume that the red giants are still in the stage of original contraction (i.e. at the beginning of their life, before nuclear transformations have even started), because then the rate of contraction necessary to explain the observed luminosities comes out too high. He has found that for a typical red giant such as ζ Aurigae *K* (with a mass equal to 15 sun-masses and a radius equal to 200 sun-radii) the radius would have to change one per cent per year and therefore the luminosity 0·2 *m* per decade ; this is contrary to observation, which gives an upper limit of about 0·005 *m* per decade. However, even more conclusive evidence against the view that the energy production of the giants comes from gravitational contraction is derived from observations on pulsating red giants, which show that the radius remains constant within a much narrower margin.

These considerations have led Gamow [16] back to the conclusion that the sources of energy in the red giants come from nuclear reactions which go faster and at a lower temperature than the reactions responsible for the energy production in the main-sequence stars. The only reactions which could fulfil these conditions are the ones involving deuterium, lithium, beryllium and boron. However, if any of these nuclei interact with protons, they are consumed

permanently, as has been pointed out already. For the reaction of a proton with Li, Be or B leads practically always to the emission of an α-particle, leaving a residual nucleus lighter than the one originally present. This process will be repeated until only helium is left, which will be the case after one or at most two reactions. Deuterium goes over directly into ^{3}He by proton capture, and this in turn transforms finally into ^{4}He. For none of these light nuclei do we get a cyclical reaction as for carbon, so that their supply will not be replenished. Therefore, if these nuclei are responsible for the energy production in giants, it must be concluded that the red giants are very young stars, created long after the separation of the great nebulae (cf. § 7).

An important conclusion which permits a test of Gamow's hypothesis is that the red giants must form, in the *L–R* diagram, several bands running parallel to the main sequence. Gamow and Teller [17] have calculated the positions of these bands, along which the stars with different nuclear reactions should be located. They then plot the pulsating red giants on the *R–L* diagram because they form a one-parameter sequence (all their characteristics are definite functions of the vibration period). They find that there are definite concentrations of stars near the places where different nuclear reactions of light elements, calculated entirely without adjustable parameters, become predominant. Thus the short-period group of cluster variables contains the stars in which the transformation of ^{10}B seems to be the most important reaction. The second group is not quite so definite; it contains the Cepheid variables, and the broad and irregular distribution is interpreted as due to the overlapping of three separate groups corresponding to the reactions of Li, Be and ^{11}B. Finally there is a third group of long-period or "Mira variables", which, according to Gamow and Teller, should be producing energy by means of the deuterium reactions (probably D+D rather than D+H). This seems to be a promising approach, and while the theory is still on a less sure footing than that of the main-sequence stars, it seems possible that thermo-nuclear processes can explain the large energy production in the red giants.

It might be remarked that it is not possible, as has been suggested on several occasions [9], [16], to explain the energy production of the red giants as arising from the formation of neutron cores with the consequent release of energy; such a core would require a much smaller radius than the red giants possess, and, moreover, would not be stable under the conditions of pressure and density prevailing in these stars.

§ 6. ENERGY PRODUCTION IN THE WHITE DWARFS

As we have already indicated, the white dwarf stars differ from "normal" stars of the same mass in two fundamental aspects: they have extremely low luminosities and exceedingly high densities (as a result of their small radii) as compared with the "normal" stars.

As has been pointed out in § 3, the high density of a white dwarf star leads to degeneracy of the electron gas throughout a considerable portion of its interior. On the other hand, the high surface temperature indicates that "normal" perfect gas conditions prevail in the outer shell of the star. These

facts imply the existence of a transition region which must be carefully treated in order to find out the internal temperature distribution. Expressions have been derived [19] for the radiative and conductive opacities and the free-electron density in both the degenerate and non-degenerate regions of the star. The surface integration of the star equations for the best-observed white dwarfs indicates that at a very short distance inward from the edge of a white dwarf strongly degenerate conditions already prevail and that the temperature is rising very slowly. This permits one to neglect, further inside, the temperature-dependent terms in the gas pressure.* In this approximation, according to Chandrasekhar's theory of degenerate configurations [20], a given total mass uniquely determines the radius and the mass-density distribution of the white dwarf. Knowing the density distribution, it is a relatively simple matter to find the associated temperature distribution without making any explicit assumptions concerning the energy-production process. In Sirius B the central density is 3×10^7 g./cm^3, the central temperature 15 million degrees ; the corresponding figures for O_2 Eridani B are 1×10^6 g./cm^3 and 40 million degrees.

From the temperature-density distribution we can calculate the concentration of reactants which must be present if the observed luminosity is assumed to be due to the occurrence of nuclear processes. The low luminosities of the white dwarfs taken in conjunction with their high densities and moderately high temperatures demand extremely low hydrogen contents. If the carbon concentration is the same as in main-sequence stars—which seems plausible—the hydrogen concentration must be 10^{-8} for Sirius B and 10^{-12} for O_2 Eridani B. Even if carbon and nitrogen were completely absent, the proton-proton reaction would still occur, and thus an upper limit for the hydrogen concentration of 10^{-5} for Sirius B and 10^{-4} for O_2 Eridani B is obtained ; it can easily be seen that such low hydrogen contents could supply the energy of the white dwarfs only for a very short time. E.g., the figure of 10^{-12} for O_2 Eridani B leads to a life of only ten years, which is manifestly wrong. This leads to the conclusion that gravitational contraction is the primary source of energy of the white dwarf stars. With this source of energy it will take at least 10^8 years before a white dwarf will become a dark object.

§ 7. STELLAR EVOLUTION

We have shown above that all important nuclear reactions in stars lead to the formation of helium out of protons and electrons but not to the building up of any heavier elements. On the contrary, in red giants there seems to take place a *destruction* of heavier elements (Li, Be, B) by reactions in which protons are captured and α-particles emitted. This fact, viz. that Li, Be, B emit α-particles upon proton bombardment rather than capture the proton, prevents any effective building up of elements heavier than helium [9]. Any such elements which are now present in the stars must therefore have been present when the stars were first formed, contrary to the *Aufbau* hypothesis [21]. Weizsacker [22] has shown that the present distribution of heavy elements is compatible with

* It can be shown that this approximation changes the results only by a fraction of 1 per cent.

the assumption that matter was in thermal equilibrium at an extremely high temperature ($>10^9$ degrees) before the stars were formed.

Concerning the evolution of the stars themselves, present theories give a picture somewhat as follows : the life of a star begins with the condensation of a large amount of matter which gradually contracts, the potential energy set free by the gravitational contraction being transformed into radiation. When the star has contracted sufficiently to make the central temperature about 200,000 degrees, the deuterons in the star begin to react with each other (and with protons) ; the star is then a " super-giant " and, if variable, a Mira variable. After a period of the order of some millions of years, the deuterium will be exhausted, and the star will resume gravitational contraction which will last for a comparatively short period (10^3 to 10^5 years, perhaps). Finally, the central temperature will reach about 5×10^6 degrees and Li, Be and B will begin to react with protons. In this *giant* stage the star may stay somewhat longer than in the deuterium-consuming supergiant stage because the total abundance of Li, Be and B may be greater than that of the deuterium. During the giant stage neither the radius nor the luminosity of the star will change very much.

When the light elements D, Li, Be and B are all consumed, further gravitational contraction will occur which brings the star over to the main sequence. The rapidity of the gravitational contraction accounts for the absence of stars between the red giants and the main sequence. Stars on the boundary line between giants and the region of gravitational contraction may be especially unstable, which may account for the variable stars [17].

The gravitational contraction will finally raise the central temperature to about 20 million degrees, so that the carbon cycle sets in. This carbon cycle will supply the energy during the main part of the life of the star ($\sim10^{10}$ years for the sun). Since the thermo-nuclear reactions supplying the energy radiated by the main-sequence stars lead essentially to the transformation of hydrogen into helium without the building up of heavy elements, the hydrogen content should be taken as the chief parameter of evolution of these stars. Gamow [16] and Bethe [9] have found that as the consumption of hydrogen continues, the central temperature and luminosity steadily increase, although the rate of change of the hydrogen concentration is not constant, the H-content decreasing slowly at first, then more and more rapidly. This is very fortunate, because it would be very difficult to account for a mass-luminosity law if the hydrogen content could vary absolutely at random. As it is, the M–L relationship is interpreted as a statistical correlation, due to the fact that the chance of observing a main-sequence star near the end of its life is very small.

When the hydrogen content is almost exhausted, the star will again start contracting with steadily increasing velocity. Very soon the gravitational energy liberated through such contraction will be larger than the nuclear energy liberation. The contraction is accompanied by a further increase in luminosity. The future course of the star depends on its mass ; for sufficiently small masses (below Chandrasekhar's limit [20] of about 1·5 sun-masses) the electron gas will finally become degenerate. This will slow down the contraction, and the luminosity will decrease until ultimately a white dwarf (and still later

a black dwarf) results. However, the time required for this evolution is so long (about 10^{11} years) that it would be difficult to believe that the white dwarfs we now observe have had their origin in this manner. Such a "long" time scale would be contrary to all the other observational evidence we possess (cf. below).

An alternative viewpoint has been advanced to explain the origin of the white dwarf stars. It is based on an examination of the evolutionary possibility of the more massive stars. For these stars (with masses larger than 1·5 sun-masses) the process of gravitational contraction is not limited by electron degeneracy, so that the contraction can continue indefinitely at a rapid rate until the centrifugal forces become so large that a break-up occurs. It has been suggested [5], [16] that the fragments resulting from such explosions will quickly become white dwarfs. This theory has the advantage that the massive stars are so luminous that the thermo-nuclear processes at first responsible for the energy production consume the hydrogen present in a comparatively short time.

Another hypothesis previously entertained concerning the evolutionary behaviour of the more massive stars—that neutron cores are formed in giants after the exhaustion of the nuclear sources of energy—has been critically examined by Oppenheimer and Volkoff [24]. They have studied the gravitational equilibrium of masses of neutrons, using the equation of state for a cold Fermi gas and the refinements of general relativity. They have found that neutron cores cannot be stable against disintegration into electrons and nuclei for masses less than about 0·1 $M_{\odot}$, and that the masses of the shells of ordinary (but dense) matter surrounding the core must be small for cores much larger than the lightest ones stable. They therefore conclude that it is unlikely that static neutron cores play any great part in the evolution of the massive stars.

In conclusion, it might be noted that not only have attempts been made to gain some understanding of the evolutionary history of individual stars, but a fruitful attack has been started on the problem of the formation of the great nebulae (vast clusters of stars). In a very interesting paper [25] Gamow and Teller deduce that the nebulae are the largest assemblies of matter which can be kept together by gravitation against the dispersing effect of the random velocities of the stars. Moreover, they point out that the mutual velocities of neighbouring nebulae are of the same order as the random velocities of stars in a nebula. This supports the theory that all nebulae originated from the same very limited region of space. When the red-shift of the spectral lines of distant nebulae is used to calculate their recession velocity, it is found that the nebulae separated about $1{\cdot}8 \times 10^9$ years ago, a figure which is in excellent agreement with geological data and statistical considerations on binary stars. It seems much more likely that such an odd occurrence as our planetary system might be formed in the original highly condensed state of the universe than in the present dilute one.

REFERENCES

(1) KUIPER, G. P. *Astrophys. J.* **88**, 429 (1938).
(2) VOGT, H. *Astr. Nachr.* **226**, 301 (1926).
(3) RUSSELL, H. N. In Russell, Dugan and Stewart, *Astronomy*, **2**, 910 (1938).
(4) EDDINGTON, A. S. *Internal Constitution of the Stars* (1926).
(5) STRÖMGREN, B. *Handb. d. Astrophys.* **7**, 172 (1936).
(5 *a*) KUIPER, G. P. *Astrophys. J.* **89**, 548 (1939).
(6) KOTHARI, D. S. *Phil. Mag.* **13**, 361 (1932).
(7) ATKINSON, R. D'E. and HOUTERMANS, F. G. *Z. Phys.* **54**, 656 (1929).
(8) GAMOW, G. and TELLER, E. *Phys. Rev.* **53**, 608 (1938).
(9) BETHE, H. A. *Phys. Rev.* **55**, 434 (1939).
(10) BETHE, H. A. and CRITCHFIELD, C. L. *Phys. Rev.* **54**, 248 (1938).
(11) MARSHAK, R. E. and BETHE, H. A. *Bull. Amer. Phys. Soc.* **14**, no. 3 (1939).
(12) KUIPER, G. P. *Astrophys. J.* **88**, 472 (1938).
(13) GAMOW, G. *Astrophys. J.* **89**, 130 (1939).
(14) RUSSELL, H. N. (Private communication.)
(15) COWLING, T. G. *Mon. Not. R. Astr. Soc.* **96**, 42 (1935).
(16) GAMOW, G. *Phys. Rev.* **55**, 718 (1939).
(17) GAMOW, G. and TELLER, E. *Phys. Rev.* **55**, 791 (1939).
(18) GAMOW, G. *Z. Astrophys.* **16**, 113 (1938).
(19) MARSHAK, R. E. and BETHE, H. A. (To be published shortly.)
(20) CHANDRASEKHAR, S. *Introduction to the Theory of Stellar Structure*, chap. XI (1939).
(21) WEIZSÄCKER, C. F. VON. *Phys. Z.* **38**, 176 (1937).
(22) WEIZSÄCKER, C. F. VON. *Phys. Z.* **39**, 633 (1938).
(23) WHIPPLE, F. L. *Proc. Nat. Acad. Sci., Wash.*, **25**, 118 (1939).
(24) OPPENHEIMER J. R. and VOLKOFF, G. M. *Phys. Rev.* **55**, 374 (1939).
(25) GAMOW, G. and TELLER, E. *Phys. Rev.* **55**, 654 (1939).

INDUCED RADIOACTIVITY

By HAROLD J. WALKE, M.Sc., Ph.D., F.Inst.P.
Exhibition of 1851 Senior Student
George Holt Physics Laboratory, The University, Liverpool *

§ 1. INTRODUCTION

During the period intervening between the discovery of induced radioactivity by Curie and Joliot [1] in 1934 and the present date, extensive investigations in this field have been carried out in many laboratories. In consequence a very great number of new radioactive isotopes have been discovered, and the properties of some of them have been studied in detail. The results which have now accumulated are already so numerous that it would be difficult to give a complete summary of them in an article such as this. As, however, several of the more recent investigations have brought to light information which has clarified certain aspects of the subject, this report will be limited to considerations of such experiments.

For convenience of reference the article and the corresponding list of publications will be divided into sections. Wherever possible an attempt will be made to preserve continuity between succeeding sections.

No list of the extensive results of induced radioactivity will be appended as Livingood and Seaborg [2] have recently published a detailed and complete list of the present data which is almost unique in that an attempt has been made to give a critical review of the published work. This extensive and critical summary is to be very highly recommended.

§ 2. THE RANGE-ENERGY RELATION FOR β-RAYS

Recently attention has once more been focussed on those experimental methods of studying the process of β-disintegration which are possible with feebly active materials. Of these the absorption method for the determination of the upper limit of continuous β-ray spectra has the advantage of being rapid and experimentally simple. With strong sources, and particularly when no appreciable γ-radiation is present, a quite definite absorption limit is found which is roughly identified with the particles of maximum energy. When sufficiently accurate measurements are made the same simple method can be applied even in the presence of strong γ-radiation from the source.

The usual procedure in determining the upper limit from the absorption curve has been to plot the relative ionization or number of counts against the thickness of absorber. The limit is then taken as the range corresponding to the point at which the curve becomes flat as estimated by the eye. With strong sources subtending a small solid angle at the detector the uncertainty introduced by visual estimation may not be large, but the method must always be subjective, and no reliable estimate of the error involved can be obtained. In cases

* Dr. Walke's untimely death on December 21st, at the age of 28, will be deeply regretted by all who knew either the man or his work.

where the weakness of the source introduces large statistical errors the visual method is much less satisfactory. In general the visual estimate of the upper limit from absorption curves obtained with weak sources is too low.

The recent investigations have produced two developments of the absorption method. The first has resulted in the evolution of methods whereby the best use of the data obtained with the weakest sources may be made, and the second whereby the absorption curve may be used to provide further information regarding the energy spectrum.

Widdowson and Champion [3] have attempted an empirical solution of the first problem. Previously, the maximum energy of the β-particles of a continuous spectrum was deduced from the absorption limit by using Feather's empirical relation

$$R=0{\cdot}511E-0{\cdot}091,$$

where R=range in gm./cm^2; E=energy in Mev.

Using the more trustworthy data available concerning maximum energies, as well as the results of their absorption measurements obtained with tube counters, Widdowson and Champion proposed the relation

$$R=0{\cdot}536E-0{\cdot}165.$$

Feather [4] has, however, developed a method which deals empirically with both problems at the same time. He has given a detailed account of a method using the β-radiation of RaE as a standard of comparison by which the maximum energy, and some indication of the form of the energy spectrum, of the β-particles from any other radio-element may be obtained. The method is applicable both to weak sources and to sources of finite thickness, and is designed to yield the maximum amount of information under such conditions.

Feather has recalculated the constants of the range-energy relation with the following result:

$$R=0{\cdot}543E-0{\cdot}160.$$

Although the shapes of the curves obtained are different if ionization chambers are used instead of counters, the new method of Feather is still applicable, and consistent results can be obtained. Moreover, the fixed procedure suggested is valuable since all data obtained using this method and range-energy relation can be directly compared. It is therefore recommended as a standard absorption method.

§ 3. THE CONTINUOUS β-RAY SPECTRUM

Interest in the continuous spectrum of disintegration electrons emitted by β-radioactive bodies was increased by the discovery of induced radioactivity and the difficulty which has been experienced of incorporating an account of the phenomenon in the theory of the nucleus. Attention has been mainly focussed on two features of the spectrum, namely, the high-energy limit, the accurate measurement of which yields the total change in nuclear energy associated with the β-disintegration, and the shape of the energy-distribution curve, which is discriminative in theories of the β-ray disintegration process.

The original theory of β-decay proposed by Fermi [5] made a definite prediction as to the energy distribution of the disintegration electrons emitted by a β-radioactive isotope. It was soon found, however, that the experimentally observed distribution did not agree with the theoretical shape, for the number of low-energy electrons detected was considerably greater than the number predicted. The theory was then modified by Uhlenbeck and Konopinski [6], who introduced into the function governing the shape of the distribution a factor proportional to the square of the momentum of the emitted neutrino. This modified theory appeared at first to be in agreement with the experimental results.

Two forms of apparatus have been principally used for investigating the shape of the continuous spectrum, the Wilson expansion chamber operating in a uniform magnetic field and the magnetic spectrograph. Paxton's investigation [7] of the scattering of electrons through small angles shows that the spurious curvatures resulting distort considerably the spectrum obtained when using the expansion chamber. In consequence, most recent measurements of β-ray energies have been made using well-designed magnetic spectrographs in which scattering from internal surfaces has been minimized or eliminated.

In this way a large number of investigations of the spectrum of RaE have been made, and recently concordant values of the upper limit of the spectrum have been obtained. However, the main factors influencing the observed energy distribution (namely, absorption in the window of the recording counter and scattering from the material of the mounting of the source) have only recently been carefully investigated and, in consequence, there is good agreement only among the most accurate work. Two investigations in which similar methods were used, and in which great care has been taken to eliminate all sources of error even at the lowest energies, have been undertaken by Flammersfeld [8] and Neary [9]. Neary's work is unique in that the complete spectrum was obtained without the necessity of applying corrections of any kind.

The spectrograph used by Neary is in the form of a rectangular aluminium box coated with colloidal graphite. The focussing slit was arranged so that secondary electron production at its edges was reduced to a minimum and a system of short diaphragms with bevelled edges was arranged around the edge walls, pointing towards the source. The effect of these diaphragms at low energies is practically to eliminate scattered or secondary electrons. Separate experiments showed indeed that the stray scattering was actually negligible. To record the particles, Geiger-Müller counters were used with axes along the magnetic field so that the entering particles approached the central wire and region of maximum efficiency. These counters were used with very thin windows (down to 0·01 mg./cm^2) in which there was negligible absorption even for the slowest β-rays measured. They gave full counting efficiency for all energies and the circuit used gave a *perfect* counting-characteristic over a range of 60 volts. The scale-of-two recording sets used gave no loss of counts even at the highest counting rates.

Both Neary and Flammersfeld and, less thoroughly, Martin and Townsend [10] and Lawson [11] and Tyler [12], have carried out experiments to determine the

effect of experimental conditions on the distribution curve obtained with β-radioactive substances. It has been found that the effect of scattering from the source mounting and absorption in the counter window can alter the energy distribution very considerably, and in a quite unpredictable manner. Flammersfeld, in fact, has shown that even aluminium foil 0·5 μ in thickness appreciably scatters electrons of energy less than ~15 kev. The complete scattering phenomenon is complicated, and its effect on the shape of the spectrum is much greater than has been previously supposed. Such effects were also observed by Lawson and Tyler, but their thinnest sources were still sufficiently thick, as judged by the work of Neary and Flammersfeld, for scattering to have been quite appreciable in the low-energy region, and their results are, therefore, not accurate below ~300,000 ev.

Flammersfeld and Neary, using RaE deposited on 0·5 μ mica and 0·5 μ aluminium mounted in light aluminium frames (so that there was no heavy material near the source), have made measurements on electrons of energies

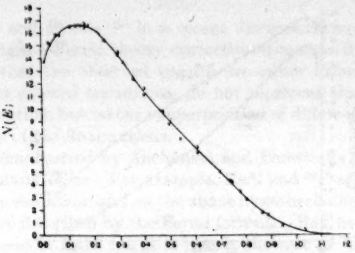


Figure 1.

as low as 5 kev. and 13 kev. respectively. It is therefore probable that the distribution of RaE is now known quite accurately down to 20 kev., for Neary in particular has shown that, in his experiments, scattering and absorption were negligible even at the lowest energies. The curve shown in figure was obtained without the necessity for any corrections or allowances for reflection, absorption, etc. The interesting point about the curve is that it has a maximum at about 150 kev. and decreases at lower energies. Neary's data on this low-energy part of the curve are very extensive, the probable errors for points even at very low energies being only of the order of 2 per cent. The end-point was not searched for carefully, in view of the concordant values already existing. Flammersfeld's value, however, is in good agreement with the better determinations, e.g. those of Langer and Whittaker [13], Lyman [14], etc., the curve approaching the end-point 1·17 Mev.

The validity of the Fermi theory of β-ray emission can be tested by plotting $\sqrt{(N/f)}$ against the energy of the particles, where N is the number of particles

emitted in a given momentum interval and f is a function which can be written approximately as

$$f=\eta^2\left(\frac{2\pi y}{1-e^{-2\pi y}}\right),$$

where $y=\dfrac{Z(1+\eta^2)^{\frac{1}{2}}}{137\eta}$ and η is the momentum in units mc.

According to the Fermi theory, the resulting curve should be linear, with an intercept on the energy axis equal to the upper limit of the spectrum. Similarly a plot of $(N/f)^{\frac{1}{2}}$ against energy would be linear with the same intercept if the Uhlenbeck-Konopinski modification were validated by experiment.

The Fermi and Uhlenbeck-Konopinski plots for RaE (Neary) are exhibited in figure 2. (The curves by Flammersfeld are essentially similar.) They show the now well-known characteristics, i.e. the Uhlenbeck-Konopinski plot is almost a straight line until approaching the upper limit, the Fermi plot being curved, but approximately linear near the upper limit, and approaching the

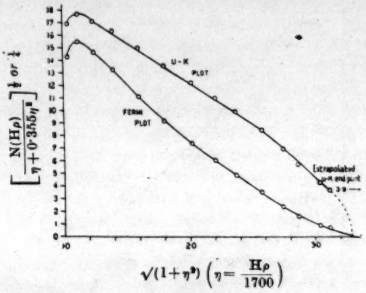


Figure 2.

experimental end-point. As in all cases so far accurately investigated, the Uhlenbeck-Konopinski plot deviates markedly from linearity in the upper energy regions ; it must be inferred that this theory is not satisfactory, for this part of the spectrum is most accurately known.

One possible explanation of the excess of particles in the low-energy region of the Fermi plot is the possibility of a finite mass for the neutrino. This mass can be determined from the extrapolated Uhlenbeck-Konopinski end-point and the observed upper limit. The value thus obtained by Neary is $0{\cdot}57\,m_0$, where m_0 is the mass of the electron. The agreement between the observed end-point of the ^{13}N positron spectrum and the mass difference between ^{13}N and ^{13}C [15] shows, however, that the neutrino mass is $0{\cdot}0\pm0{\cdot}2\,m_0$, and, moreover, the observed curves for RaE are not in conformity with the simple corrections for finite neutrino mass. *It thus appears that neither the original theory of Fermi nor the modification by Uhlenbeck and Konopinski can adequately predict the complete spectrum of disintegration electrons from a radioactive body.*

Other work using ^{32}P [11], ^{24}Na [11] and ^{64}Cu [12] has shown that in these cases the Fermi theory fits the *higher* energy portion of the spectrum well. Tyler [12] has shown, moreover, that with thick sources the scattering etc. has the effect of distorting the spectrum so as to allow a closer fit with the Uhlenbeck-Konopinski modification. However, these workers [11], [12] were using essentially thick sources, so the low-energy regions of the spectra obtained are probably considerably distorted.

Since the Fermi theory accounts satisfactorily for the observed distributions near the high-energy limit, and the Uhlenbeck-Konopinski theory for a range somewhat removed from this limit, it is tempting to employ a linear combination of the two distributions to represent the whole range of results. The serious departure from the Uhlenbeck-Konopinski distribution for RaE near the low-energy end, however, would require the addition of a further term in which still higher powers of the neutrino momentum would be involved. This procedure would, therefore, lose most of its attractiveness since, with three arbitrary constants, the basic theories would no longer play a significant rôle.

Bethe, Hoyle and Peierls [16] in a recent discussion have investigated the view that the original Fermi theory correctly represents the elementary law of β-decay, but that the observed spectra are either forbidden or, in so far as they belong to allowed transitions, do not represent the effect of a single nuclear transformation, but rather a superposition of different spectra belonging to different levels of the final nucleus.

However, as emphasized by Alichanian and Berestezky [17], this viewpoint is untenable in many cases. For example, RaE and ^{30}P are not accompanied by any primary γ-radiation, and on the above hypothesis the β-spectra of these elements would be described by the Fermi formula. But, as shown above, this is not so in the case of RaE, nor is it true in the case of ^{30}P, the spectrum of which shows a sharp contradiction to the Fermi spectrum and a better agreement with the Uhlenbeck-Konopinski spectrum. Exactly similar results have been obtained with ThC and ThB, both of which are disintegrations of the "allowed" type. Alichanian and Berestezky state that the suggestion that disagreement with the Fermi theory may be explained in certain cases by the fact that these transitions are forbidden is unsatisfactory, for the complete symmetry of the electron-neutrino interaction term of the Fermi theory cannot explain the experimentally observed asymmetry of the β-spectrum.

It seems then not unlikely that some fundamental change may be needed to relate the theory of β-decay to the results obtained experimentally.

§ 4. NUCLEAR K-ELECTRON CAPTURE

It follows as a general consequence of current theory of nuclear β-transformation that a nucleus which is capable of emitting a positron has a certain probability of making the same transformation by capturing an electron from the extranuclear structure, the atomic electron most likely to be thus captured being one of the K-electrons. The relative probabilities of these alternative

modes of transformation depend on certain detailed features of the theories, and, as was first shown by Møller [18], the requirements in this respect of Fermi's [5] theory of β-decay are very different from those of the modification of Uhlenbeck and Konopinski [6].

The capture of a K-electron by an atomic nucleus would be followed by the emission of K-radiation due to the filling up of the K-shell. This radiation is indeed the most characteristic external effect that would be produced, and its observation constituted the first evidence for the phenomenon. This was made by Alvarez [19], who found that titanium after bombardment by deuterons emitted a radiation with an absorption coefficient in excellent agreement with that of the K-radiation of titanium.

Strictly speaking, however, the observation of the K-radiation provides evidence only for the removal of K-electrons from the atoms concerned. In order definitely to establish the phenomenon of nuclear transformation by electron capture it must be proved, as Alvarez [20] has emphasized, that no other process is responsible for their removal.

Alvarez [20] has given a rigorous proof of the occurrence of the phenomenon in the radioactive isotope ^{67}Ga. This body was separated from the other activities induced in zinc by deuteron bombardment by means of the ether-extraction process. It was shown by " characteristic absorption " measurements, using nickel and copper foils, that ^{67}Ga emits Zn Kα and Kβ radiation, and it was proved that this nucleus may change into ^{67}Zn without the emission of a positron. The ^{67}Ga nucleus decays by capturing one of its K-electrons. In the case of ^{67}Ga, however, γ-radiation and internal-conversion electrons are also emitted, and while there is no doubt that this isotope decays by K-electron capture, a more clear-cut example is given by ^{47}V.

Walke, Williams and Evans [21] have established that this isotope, of half-life 600 ± 50 days, decays *only* by K-electron capture. The active body was separated chemically from titanium which had been activated by deuterons and was found to emit only x-radiation. No ionizing radiations nor any γ-radiation have been detected, yet the vanadium precipitate was found to emit Ti Kα radiation. No hypothesis other than decay by nuclear K-electron capture will explain this result. Suppose, for example, that the K-electron of such a vanadium atom is removed by electronic or quantum radiation from other vanadium atoms. In this case the K-radiation emitted would be that of *vanadium*. This would be so also if the K-electron were removed by internal conversion of γ-radiation from the nucleus of the same atom. If it were removed by a *disintegration electron* from the nucleus of the same atom, the K-radiation emitted would be that of *chromium*. The fact that the K-radiation is that of titanium rules out all these processes. There remains the possibility that the vanadium nucleus emits a slow positron, which either removes a K-electron in leaving the atom, or leaves the product nucleus in an excited state, the γ-ray subsequently emitted being highly internally converted. The K-radiation emitted would in this case be that of titanium. Owing to the annihilation of the positrons, however, two quanta of 0·5-Mev. energy would also be emitted for every nuclear transformation. This gives an intensity of γ-radiation at least

1000 times greater than that observed. Thus no hypothesis other than that of nuclear transformation by electron capture will explain all the observations.

The fact that the activity of ^{47}V is not accompanied by the emission of any detectable radiation other than K-radiation shows that the electron is captured into the ground state of the final nucleus. But the circumstance has additional interest. Usually when a system drops from one state to another the energy difference is emitted in the form of radiation, and the absence of any such radiation in the transition represented by the above electron-capture process would appear to be a very striking exception to this rule. The result is explained on the current theory of β-transformation by the emission of a neutrino which takes away the energy difference between the initial and final states of the atom.

The relative probability of electron capture and positron emission by a given nucleus is a matter of considerable interest, particularly as the current theories of β-transformation, viz., that of Fermi [5] and its modification by Uhlenbeck and Konopinski [6], have appreciably different requirements in this respect. There are, however, very little data available on this point.

Walke, Williams and Evans [21] have investigated these relative probabilities for the case of ^{48}V (half-life 16 days, upper limit of positron spectrum 1·0 Mev.). For this case, according to calculations carried out by Møller [18], the ratio, $R=F_K/F_P=$(probability of K-capture)/(probability of positron emission) $=$ 0·10 in Fermi's theory; $R=1·0$ in the Uhlenbeck-Konopinski modification. They actually obtained a value $0·4\pm(0·15)$.

This result is of the same order of magnitude as the theoretical values, being actually between the value required by Fermi's theory and that required by the theory of Uhlenbeck and Konopinski. The theoretical computations assume the transition concerned, viz. $^{48}V\rightarrow{}^{48}Ti$, to be an allowed one. The long half-life (16 days) associated with the positrons of energy 1·0 Mev. indicates that the transition is actually strongly forbidden, and it is to be expected on general grounds that R would be greater for such transitions. As the observed value is much less than that required by the Uhlenbeck-Konopinski theory even for allowed transitions, the above result must be regarded as more favourable to Fermi's theory.

It is probable that decay by K-electron capture has been established with the following radioactive nuclei :—

^{7}Be [22], ^{40}K [23], ^{41}Ca [24], ^{46}Sc [21], ^{47}V [21], ^{48}V [21], ^{51}Cr [34], ^{52}Mn [25], ^{54}Mn [26], ^{56}Co [27], ^{65}Zn [28], ^{66}Ga [29], ^{67}Ga [20], two isotopes of element 43 [30], ^{106}Ag [31], (?) ^{107} or ^{109}Cd [32], ^{121}Te [33], ^{152}Eu [34], (?) ^{180}Ta [35], ^{197}Hg [36].

§ 5. NUCLEAR ISOMERISM

It was first suggested by Soddy [37] in 1917 that there might exist nuclei with the same atomic and mass number, which nevertheless had different radioactive properties. Such nuclei, which he named nuclear isomers, were then unknown, but evidence in favour of this hypothesis of nuclear isomerism was

first obtained by Hahn [38] in 1921. Hahn made a careful study of the chemical and radioactive properties of UZ and proved that it is isomeric with UX_2. Recent work by Feather and Bretscher [39] has confirmed this deduction and has elucidated the probable genetic relationship between these two substances.

A natural consequence of the search for radioactive isotopes of the stable elements has been the discovery that nuclear isomerism is by no means an exceptional phenomenon. Many isomeric pairs are already known, the first certain example provided by induced radioactivity being ^{80}Br, which was established by studies of the radioactivities induced in bromine by slow and fast neutrons, deuterons and γ-rays. The half-lives of the two isomers of ^{80}Br are 18 min. and 4·5 hours respectively.

Weizsäcker [40] has suggested that such isomeric pairs are to be explained by assuming that the lowest excited state of the nucleus has an angular momentum differing by several units from that of the ground state. Selection rules can then be invoked to weaken considerably the probability per unit time of the transition from the upper to the ground state of the nucleus. The excited state is thus metastable against γ-decay and can exist for appreciable times before radiating.

On the basis of this picture, Weizsäcker has estimated the lifetimes for states of various angular momenta and energies. For large spin changes and small energy differences these lifetimes may be very long.

In consequence, the study of the γ-radiation eventually emitted in the transition from one isomeric state to another becomes a matter of importance. But since in many cases the nucleus will also be unstable by β-decay, the upper state, corresponding to one of the isomeric forms, may be destroyed by a competing β-transition. The β-γ branching ratio in such cases will depend on the decay constants for the two modes of decay.

The early attempts to observe the γ-radiation were unsuccessful. This is explicable, however, as was first pointed out by Pontecorvo [41], if the γ-ray due to the transitions between isomeric states is highly internally converted.

Hebb and Uhlenbeck [42] have therefore investigated the effect of internal conversion of the γ-radiation emitted, using Bethe's estimates for the half-lives of the metastable states. It was found that for large changes in the angular momentum of the nucleus and for small energy differences between the isomeric states, the internal conversion coefficient is approximately unity. Hence one would expect the transformation of one isomer into another to be accompanied mainly by the emission of conversion electrons and not by γ-quanta as assumed by Weizsäcker.

This effect was first noticed in the case of the isomers of ^{104}Rh by Pontecorvo [43], who observed a group of low-energy electrons which he interpreted as an electron line due to the internal conversion of the γ-ray emitted in the transition from the metastable level to the ground state of the nucleus. Recent experiments have now qualitatively confirmed Hebb and Uhlenbeck's theory, and there can be little doubt that in the majority of cases one isomer

transforms into the other by emitting γ-radiation which is highly internally converted.

It is of particular interest to have methods of investigating possible genetic relationships between the isomeric states of the same nucleus. This can be readily accomplished if the lower level has a longer life than the upper. But for the opposite case the problem is much more difficult, because a transient equilibrium is soon reached and the only apparent period is that of the parent. In this connexion a brilliant method, whereby an actual separation of genetically related isomers can be accomplished, has been developed by Segrè, Halford and Seaborg [44]. The principle of the method is as follows.

Consider an element with isomeric nuclei which can give compounds suitable for the application of the Szilard-Chalmers [45] method of concentration. Suppose the isomer in the upper state decays to the lower state; then, if the γ-ray energy is high enough, the corresponding recoil can be sufficient to eject the decayed atom from the compound. The transformation product can then be separated as in the Szilard-Chalmers method. Moreover, if the energy of the γ-ray is so low that the recoil would not be sufficient to break the chemical bond, then if the γ-ray is highly converted, the recoil from the corresponding conversion electron may be so much higher that recoil energies comparable with chemical bond energies are obtained.

Even when the recoil energy is not sufficient to break the bond, the method does not necessarily fail, because the energy given to the molecule by the ionization of the K or L shells may suffice to create an activated condition within the molecule such that it can undergo chemical reaction with some other kind of molecule deliberately introduced for the purpose, or simply decompose.

In the case of ^{80}Br it was thus found that the recoil from the conversion electron is sufficient to break a chemical bond in the BrO_3^- anion, or to activate the C-Br bond in tertiary butyl bromide and to facilitate the hydrolysis of the compound. The experiments of Segrè, Halford and Seaborg [44] and of De Vault and Libby [46], in which a *complete* separation of the isomers was obtained, have shown conclusively that the period of 4·5 hours is to be associated with a metastable state of ^{80}Br, the ground state decaying by β-ray emission with a half-life of 18 min. The transformation product due to the decay of the upper state has thus been found to decay with the 18-min. half-life, while the activity of the parent (after freeing it from its product) increases with time as the equilibrium concentration of the 18-min. isomer is restored by the decay of the parent. It has been concluded that practically all of the nuclei in the metastable state decay by falling to the ground state from which the observed disintegration electrons are emitted.

A large separation of the isomers of ^{80}Br has also been obtained by a method due to Le Roux, Lu and Sugden [47].

That the γ-ray due to the transition to the ground state is highly internally converted has been shown by two methods. In the first, due to Valley and McCreary [48] and Siday [49], it has been established that electron lines are emitted by the 4·5-hour isomer. Thus Valley and McCreary have examined the

low-energy region of the spectrum and have photographed with a magnetic spectrograph two strong lines at 44 ± 1 kev. and $33{\cdot}5\pm1$ kev. respectively due to K and L internal conversion in bromine of a γ-ray of energy 45 kev. A much weaker line at $21{\cdot}5\pm2$ kev. has not definitely been identified.

Siday has investigated the β-ray spectrum of the same isomer with a Wilson chamber and has observed the homogeneous energy groups superimposed on a background due to the continuous β-ray spectrum. It was observed that approximately half the conversion electrons are accompanied by short tracks starting from the same atom and due to the Auger conversion of the Kα x-rays of bromine. These experiments show that the γ-rays have an energy of about 43 kev., in good agreement with the results of Valley and McCreary.

As a result of the internal conversion, the characteristic Kα-radiation of bromine is also emitted by the 4·5-hour isomer of ^{80}Br. These x-rays have been identified as Br Kα radiation by Abelson [50] using a bent crystal spectrograph, and his results have been confirmed by characteristic absorption measurements by Roussinow and Yusephovich [51], who have, in addition, shown that the x-rays decay with a half-life of 4·5 hours.

Table 1. *List of established pairs of isomeric nuclei*

Radioactive isotope	Half-lives	Metastable form	References
^{44}Sc	4 hrs. ; 52 hrs.	52 hrs.	A
^{51}Ti	2·8 min. ; 72 days	2·8 min. (?)	B
^{52}Mn	21 min. ; 6·5 days	—	C
^{69}Zn	57 min. ; 13·8 hrs.	13·8 hrs.	D
^{80}Br	18 min. ; 4·5 hrs.	4·5 hrs.	E
^{104}Rh	44 sec. ; 4·2 min.	4·2 min.	F
^{106}Ag	25 min. ; 8 days	8 days	G
$^{114\text{or}116}$In	13 sec. ; 54 min.	—	H
^{127}Te	10 hrs. ; 90 days	90 days	I
^{129}Te	70 min. ; 30 days	30 days	J
^{131}Te	25 min. ; 30 hrs.	30 hrs.	K
^{134}Cs	1·5 hr. ; $\sim$1 year	—	L
^{177}Lu	4 hrs. ; 6 days	—	M
^{190}Ta	14–21 min ; 8·2 hrs.	—	N
^{197}Pt	18 hrs. ; 3·3 days	—	O
^{198}Au	13 hrs. ; 4–5 days	—	P
^{234}Pa	UX_2 ; UZ.	UX_2	Q

These results suggest that the metastable level of ^{80}Br is at 45 kev. above the ground state, from which the lower isomer is formed by a γ-transition, the γ-radiation being highly internally converted. They therefore offer at least a qualitative confirmation of the hypotheses of Weizsäcker and Hebb and Uhlenbeck.

Similar results on the actual " chemical " separation of isomers have been obtained with the radioactive isotopes of tellurium. Thus it was established

by Seaborg, Livingood and Kennedy [52] that the radioactive iodine isotope ^{131}I can be grown from two forms of ^{131}Te with half-lives of 25 minutes and 1·2 days respectively. Moreover, it was possible to separate the 25-minute body at intervals until the 1·2-day activity had completely decayed. Actually three pairs of isomers have been definitely established, ^{127}Te, ^{129}Te and ^{131}Te, the corresponding half-lives being ^{127}Te : 90 days and 10 hours ; ^{129}Te : 30 days and 70 minutes ; ^{131}Te : 1·2 days and 25 minutes. It is remarkable that all three isomeric pairs are "chemically separable", the short period growing from the long-lived activity in each case, suggesting that the metastable levels have the long half-lives.

Feather and Bretscher [30] have pointed out that if isomerism is to be explained in terms of close nuclear states differing considerably in nuclear spin, it is to be expected that the disintegration of one or the other isomeric nucleus will frequently be followed by the emission of two or more γ-ray quanta in succession.

Feather and Dunworth [53] therefore carried out experiments with counters and coincidence-counting circuits with resolving times as low as 10^{-6} sec. to explore this possibility in a number of cases, and particularly to test for such cascade γ-rays following the β-disintegration of UZ. They have obtained definite evidence for the latter effect, and have studied the isomerism of ^{106}Ag with positive results. In the case of the isomer of ^{106}Ag of half-life 8 days, it was established that four or five quanta in succession are emitted following the transformation of each nucleus. In other cases, however, notably in element 43 and ^{69}Zn, such coincidence experiments have yielded negative results.

§ 6. INDUCED RADIOACTIVITY BY NUCLEAR EXCITATION

The possibility of the occurrence of *metastable isomers of stable isotopes* was first suggested by an experiment by Dodé and Pontecorvo [54], who observed a weak activity, chemically identified with cadmium, of half-life 50 minutes when cadmium is bombarded with fast neutrons. The soft radiation emitted they interpreted as proceeding from a metastable state of one of the stable cadmium isotopes which was excited by a reaction (Cd, n, n). More recently good experimental evidence has been obtained with indium and other elements which has established that such metastable isomers of some of the stable isotopes occur not infrequently. Thus when indium is bombarded with fast neutrons from a radon-beryllium source a radioactive indium isotope of half-life 4·1 hours is produced which emits slow negative electrons. Goldhaber, Hill and Szilard [55] have shown that this activity is brought about by nuclear excitation resulting in the formation of a metastable isomer of the stable ^{115}In indicated thus, $^{115}In^{*}$.

This conclusion was obtained by studying the radioactivities produced in indium by neutrons of different energy distributions, together with the successive transformations produced in cadmium by fast and slow neutrons. It was found that the body with the 4·1-hour period can be formed in indium by neutrons of energy 2·5 Mev., but not noticeably by slow neutrons, showing that it cannot be associated with an isomer of ^{116}In. Moreover, the strong

activation by 2·5-Mev. neutrons cannot be attributed to a reaction of the type $(n, 2n)$ from either ^{113}In or ^{115}In, for such a reaction requires very fast neutrons. The experiments, therefore, exclude an indium isotope of mass number 112, 114 or 116 as the carrier of the 4·1-hour activity.

In order to explain these results, Goldhaber, Hill and Szilard assumed that the ^{115}In nucleus may be excited by neutron impact and left in a metastable state ^{115}In*. They suggested that this metastable level might be radioactive and transform either by β-disintegration into its stable isobar ^{115}Sn, or by the emission of γ-rays and internal conversion electrons into the stable ground state ^{115}In. Recent work by Cork and Lawson [56] has shown, however, that the electron emission is clearly resolved in the magnetic spectrograph into the K and L conversion electron lines of a γ-ray of energy 0·336 Mev. They observed no " disintegration " electrons. It appears, therefore, that the metastable state decays by γ-transitions to the ground state.

^{115}In* has also been isolated from cadmium by activation with fast and slow neutrons and deuterons. By successive separations of indium from the activated cadmium it has been established that the ^{115}In* grows from a parent ^{115}Cd which has a half-life of 56 hours. The reactions investigated are :

(*a*) Fast neutrons : $^{116}\mathrm{Cd} + {}^{1}n \rightarrow {}^{115}\mathrm{Cd} + 2{}^{1}n$
$^{115}\mathrm{Cd} \rightarrow \beta \rightarrow {}^{115}\mathrm{In}^*$

(*b*) Slow neutrons : $^{114}\mathrm{Cd} + {}^{1}n \rightarrow {}^{115}\mathrm{Cd} + \gamma$
$^{115}\mathrm{Cd} \rightarrow \beta \rightarrow {}^{115}\mathrm{In}^*$

It is probable that the existence of a metastable state of ^{115}In is connected with its high nuclear spin (9/2). If the isomeric form is due to a state which has a small spin it will be difficult to reach the metastable state from the ground state by a direct transition. In general, no matter how produced, a higher level of ^{115}In will have to be excited from which the metastable state can be reached by one or more spontaneous transitions. The apparent threshold for the formation of ^{115}In* will thus in general be considerably above the level of the metastable state.

It has also been found possible to excite this metastable state by means of high-speed protons and α-particles. Thus Barnes and Aradine [57] have observed the 4·1-hour period by bombarding indium with 6·7-Mev. protons, and Lark-Horovitz, Risser and Smith [58] by activating indium with 16-Mev. α-particles. In each case the activity was chemically inseparable from the indium, and control experiments have excluded the possibility of the effects being due to stray neutrons. It is therefore established that the metastable ^{115}In* can be formed in the reactions ^{115}In (^{1}H, ^{1}H) ^{115}In*; ^{115}In (^{4}He, ^{4}He) ^{115}In*.

It has also been found that the same metastable state can be excited when indium is irradiated with X-rays. Thus Pontecorvo and Lazard [59] have observed the 4·1-hour period using an impulse generator giving a maximum X-ray energy of 1·85 Mev. In addition, Collins, Waldman, Stubblefield and Goldhaber [60] have obtained a yield curve by measuring the activity as a function of the energy of the X-rays. The threshold has been observed at

~1·15 Mev., and an abrupt change of slope at 1·5 Mev. indicates that there are probably two activation levels from which the metastable $^{115}In^*$ can be produced by radiative transitions.

Recently Barnes [61] has detected a similar "γ-radioactivity" of half-life 105 minutes, which is due to a metastable isomer of ^{113}In. In this case conversion electrons and a γ-ray of energy 0·39 Mev. have been observed. Other examples have been established with barium [62], element 43 [63] (see § 7) and ^{83}Kr [64]. The metastable form of ^{83}Kr of half-life 113 minutes can be produced by the disintegration of thorium by fast neutrons or by the bombardment of selenium with deuterons. It is formed from ^{83}Se and ^{83}Br, thus :—

$$^{83}Se \rightarrow \beta \rightarrow {}^{83}Br \rightarrow \beta \rightarrow {}^{83}Kr^* \rightarrow \gamma \rightarrow {}^{83}Kr$$
$$\text{24 min.} \qquad \text{2·5 hrs.} \qquad \text{113 min.}$$

The radiations emitted are very soft, having an energy of ~30,000 ev.

Strong samples of this form of ^{83}Kr can be obtained as an emanation from the bromine extracted from active selenium. Segrè therefore proposes to introduce it into a Paneth apparatus in order to investigate its spectrum. Since the spin of the ground state of ^{83}Kr is already known (9/2), it should be possible to measure the spin change involved in the transition $^{83}Kr^* \rightarrow {}^{83}Kr$ by hyperfine-structure measurements on the spectrum of $^{83}Kr^*$. In this way a direct test of Weizsäcker's hypothesis can be made.

§ 7. ELEMENT 43

Induced radioactivity has supplied chemistry with a powerful new instrument, for it is now possible to create by transmutation from other elements the element one wishes to study. In this way radioactive isotopes of element 43 have been produced and some of their properties studied in detail. The substance was first found by Perrier and Segrè [66] in molybdenum bombarded in the Berkeley cyclotron with deuterons of high energy.

At least five radioactive isotopes of atomic number 43 have been recognized among the products of bombardment, and two of them have half-lives long enough to permit a close investigation of the chemical behaviour of the new element.

Many of the chemical processes necessary to separate element 43 from other elements are quite straightforward, though separation from the neighbouring elements molybdenum (42) and ruthenium (44) is a more complicated matter. Emphasis has to be placed on the separation from rhenium, the upper homologue of element 43 in the periodic table, and which most closely resembles it chemically. The similarity between the properties of element 43 and rhenium is very pronounced, as, for example, in their reactions with H_2S, "nitron", etc. However, element 43 can be separated from rhenium by the method used by Geilmann and Weibke [67] for separating molybdenum and rhenium or by precipitating in 10N HCl by H_2S [68]. These processes are especially important since they are the only methods available for separating element 43 from rhenium at present. However, in induced radioactivity the

separation of element 43 from molybdenum is more important, and this can be readily carried out using oxyquinoline and H_2S [67].

Practically none of the physical constants of the new element can be determined owing to the small amounts of substance which can be produced artificially ($\sim 10^{-10}$ gm.). On the other hand, properties such as the volatility of certain compounds, partition coefficients in fractional crystallization, etc., can be determined with these minute quantities of substance, and some of these have actually been measured.

Cacciapuoti and Segrè [69] have investigated the decay of the long-lived isotopes of element 43. One activity of half-life 90 days is due mainly to slow electrons, a second emits electrons of higher energy and γ-radiation, with a half-life of 62 days. Since these activities have been found to be accompanied by molybdenum K-radiation, and by electron lines due to internal conversion of γ-rays, they may both decay mainly by K-electron capture.

Seaborg and Segrè [70] have recently reported an interesting case of isomerism during an investigation of the short-lived radioactive isotopes of element 43. Molybdenum irradiated with deuterons and neutrons showed a strong activity of half-life 67 ± 2 hours chemically inseparable from molybdenum. The active isotope is probably ^{99}Mo or ^{101}Mo, and as the emitted particles are negative electrons, the decay product must be an isotope of element 43. As this element is unlikely to have any stable isotopes [71], [72], it was natural to look for the growth of a radioactive product by making successive chemical separations from the active molybdenum. This has in fact been observed, and an isotope of element 43 of half-life $6{\cdot}6 \pm 0{\cdot}4$ hours has been separated. It was found by Segrè and Seaborg, and confirmed by Kalbfell [62], that the emitted electrons consist of lines due to K and L conversion of a γ-ray of energy 136 kv. The 6·6-hour activity is thus due to a metastable state which decays to a lower state by a γ-transition : the γ-ray is very highly internally converted.

This was confirmed by observations of the emitted x-rays, which were proved by characteristic absorption measurements to be the K-radiation of element 43. Moreover, Abelson [73] has photographed the Kα-line with a bent crystal spectrograph. This is the first time that the x-ray line of a new element synthesized artificially has been measured by the crystal method.

The product of the transition is presumably an isotope of element 43 in the ground state. No decay of this isotope to ruthenium has been detected. It may be stable or it may have a very long period of decay. The observations of Seaborg and Segrè show that its half-life must be at least 40 years.

No systematic study of the short-lived activities of element 43 has yet been made, but two other activities may be reasonably ascribed to this element. Provisional values of the half-lives are 2 hours and 2 days respectively. The latter emits molybdenum Kα-radiation in strong intensity, indicating that this period is associated with an isotope which decays into molybdenum, probably by K-electron capture. The shorter-lived activity has also been produced by bombarding molybdenum with high-speed protons, and its chemical identity with element 43 has been established.

A summary of the radioactive isotopes of element 43 is given in table 2.

Table 2. *Observed radioactivities of element 43*

(B=isotope probable ; C=one of few isotopes ; D=element certain ; E=element probable ; i.c.=internal conversion)

Mass number	Class (Seaborg)	Radiations emitted	Half-life	Energy of radiation		Reactions in which formed
				Electrons (Mev.)	γ rays (Mev.)	
96	B	$\bar{e}$ (?)	2·7 hrs.	—	—	Cb, ^{4}He, n Mo, ^{1}H, n Mo, ^{1}H, n
99 or 101	C	$\bar{e}$ (i.c.) 43 K_α	6·6 hrs.	0·116 0·133	0·18	β-decay of $^{99 \text{ or } 101}$Mo
	D	MoK_α; $\bar{e}$	90 days	—	—	Mo, ^{2}H, n
	D	MoK_α; $\bar{e}$	62 days	—	—	Mo, ^{2}H, n
	D	$\bar{e}$	110 hrs.	0·6	—	Mo, ^{1}H, n
	E	$\bar{e}$	53 min.	2·3	—	Mo, ^{1}H, n
	E	$\bar{e}$	36·5 hrs.	—	—	Mo, ^{1}H, n
	E	$\bar{e}$	01 min.	—	—	Mo, ^{1}H, n
	E	$\bar{e}$	0·5 min.	—	—	Mo, ^{1}H, n

§ 8. INDUCED RADIOACTIVITY BY HARD γ-RAYS (NUCLEAR PHOTO-DISINTEGRATION)

In 1937 it was shown by Bothe and Gentner [74] that the γ-radiation of energy ~17 Mev. emitted when lithium is bombarded with protons (reaction ^{7}Li (^{1}H, γ) ^{8}Be) can produce nuclear photo-effects in a large number of substances. In all cases so far established the effect consists in the separation of a neutron, this being demonstrated by the radioactivity of the resultant nucleus. In this way a number of new radioactive isotopes were discovered, namely, ^{63}Zn, ^{68}Ga, ^{78}Br, Mo (17 min.), ^{106}Ag, In (1·1 min.), ^{120}Sb, Te (60 min.). Moreover, as a result of these experiments it was possible to assign the correct mass numbers to some radioactive bodies whose half-lives and chemical properties were already known. In particular, in the case of ^{80}Br, nuclear isomerism was clearly demonstrated. Evidence of a confirmatory nature was obtained by Chang, Goldhaber and Sagane [75].

From the theoretical standpoint it is important to know how the results obtained depend on the wave-length of the γ-radiation, and whether there is any selective effect (which seemed probable from Bothe and Gentner's early work, for the effect was only obtained with certain atoms). As this question could only be decided with certainty by further experiments, Bothe and Gentner [76] have recently completed an investigation using as a second γ-ray source boron bombarded with protons. According to Fowler, Gaerttner and Lauritsen [77] the spectrum of the γ-rays capable of producing nuclear

photo-effects consists of a main line at 11·8 Mev and a second one, of one-seventh the intensity, at 16·6 Mev.

The activities obtained using the boron γ-rays were always weaker than those obtained using the Li+^{1}H γ-ray. For this reason Bothe and Gentner limited their attention to the elements with which they had previously obtained definite results with the Li+^{1}H γ-rays. However, those elements which had previously shown no definite effect with the lithium γ-rays were reinvestigated where it was known that the expected radioactive isotope would have such a half-life that they could not have avoided detecting it if the effect had been as pronounced as in those elements which displayed a positive effect. Actually new weak activities were only detected with oxygen, chlorine and selenium.

Careful control experiments established that the observed activities were not to any extent due to neutrons.

In table 3 is a summary of the results obtained, which shows that the relative effective cross-sections for the two γ-rays are not very different. If a nucleus does not show the effect (or only shows it weakly) with one γ-ray, the same applies to the other. The one striking exception to this is ^{80}Br (4·5 hours).

Table 3. *Relative yields of radioactive isotopes by nuclear photo-disintegration*

Active product	Half-life (min.)	Relative yield		Relative yield corrected for abundance of isotope activated	
		Li	B	Li	B
^{13}N	10·3	1·5	8	1	6
^{15}O	2·1	3	5	2	4
^{18}F	112	5	5	4	4
^{27}Si	6·7	2	8	2	6
^{30}P	3	31	22	21	15
^{34}Cl	33	6	8	5	7
^{44}Sc	240	—	—	—	—
^{62}Cu	10·5	100	100	100	100
^{63}Zn	38	30	25	40	34
^{66}Ga	60	55	66	61	73
^{70}Ga	20	13	27	23	48
? Se	17	3	8	—	—
? Se	57	5		—	—
^{78}Br	6	43	44	59	60
^{80}Br	18	52	37	71	50
^{80}Br	270	58	14	79	19
Mo ?	17	23	15	—	—
^{106}Ag	24	50	43	65	56
^{108}Ag	2·3	40	62	57	89
^{120}Sb	15	35	32	43	39
? Te	60	11	20	—	—

In the case of copper it was calculated that the mean effective cross-section per quantum of the lithium and boron γ-rays is 5×10^{-26} cm^2 and $2{\cdot}2\times10^{-26}$ cm^2 respectively.

The complete quantitative interpretation of the results is somewhat difficult, because of the uncertainty of the exact spectral composition of the two γ-groups. But the clear difference of the absolute cross-section for copper, as also the marked difference in the ratio of the relative production of the isomers of ^{80}Br, shows plainly that the *effective* energies of the two γ-rays must be materially different.

The results given in table 3 show further that in comparing the effects of a γ-ray of $\sim$12-Mev. with one of $\sim$17-Mev. energy, the cross-sections for the nuclear photo-effect alter in a ratio which does not differ widely for the various nuclei. In general, the absorption increases with increasing γ-ray energy. This regularity must be interpreted as meaning that the absorption spectrum of the nuclei is essentially continuous.

On the basis of Bohr's nuclear model the nuclear photo-effect is to be described as the production, by the absorption of the γ-quantum, of a highly excited nucleus, which then disintegrates by emitting a neutron. In the region of excitation here being considered the possible excited states are so close together that they form a continuum, so that the continuous nature of the nuclear absorption is readily understood.

A further interesting point is, that although no spectral selectivity appears, the yield with the various nuclei is so different. With some elements no definite effect at all could be observed, although the expected product is definitely known and of convenient half-life.

One cannot exclude the possibility that in addition to the emission of neutrons there may be other strongly competing possible modes of disintegration of the intermediate nucleus, such as a re-emission of the γ-ray or emission of a charged particle. Such individual peculiarities are to be expected more readily with the light nuclei. For this reason it is probably significant that it is with the elements of atomic number less than that of phosphorus that the effective cross-sections are found to be smaller.

§ 9. INDUCED RADIOACTIVITY BY THE FISSION OF URANIUM AND THORIUM NUCLEI

By bombarding uranium with neutrons, Fermi [78] and his co-workers detected four radioactive substances, to two of which atomic numbers greater than 92 were ascribed. Further investigations by Meitner, Hahn and Strassmann [79] demonstrated the existence of more than nine radioactive periods, six of which were assigned to trans-uranic elements. In order to explain the chemical behaviour and relationships between these bodies, it was found necessary to assume the existence of several nuclear isomers, and, moreover, at least two parallel isomeric radioactive series.

In making chemical assignments, it was always assumed that these radioactive bodies had atomic numbers near to that of uranium ($Z=92$), for all

experience with the lighter elements had shown that only protons or alpha-particles are emitted when those elements are disintegrated by neutrons. Thus a body with properties similar to those of osmium was assumed to be eka-osmium ($Z=94$) rather than osmium or ruthenium ($Z=76$ or 44). When Hahn, Meitner and Strassman observed a series of bodies with some chemical properties similar to the platinum group, they therefore classified them as higher homologues of rhenium, osmium, iridium and platinum with atomic numbers 93, 94, 95 and 96 obtained by successive β-transformations of short life from uranium.

The conditions of formation of these "trans-uranic" bodies were quite unexpected, for they were produced both by the action of slow and of fast neutrons and about as efficiently by the one as by the other.

A great difficulty arose when Curie and Savitch [80] showed that there is formed in irradiated uranium a radio-element having the chemical properties of a rare earth. This body could be chemically separated from actinium and was produced both by slow and fast neutrons under the same conditions as the trans-uranic elements. In the absence of other evidence it was therefore assumed that this body must also be trans-uranic in origin, a conclusion which was not very satisfying as it was not at all easy to assign an atomic number to this activity.

While confirming and extending this work, Hahn and Strassmann [81] found, in addition, a group of at least three radioactive bodies formed in uranium under neutron bombardment chemically similar to barium and, therefore, presumably isotopic with radium. Further investigations, however, showed that it was impossible to separate these bodies from barium (although mesothorium, an isotope of radium, was readily separated in the same experiment). Hahn and Strassmann were thus forced to conclude "*that isotopes of barium ($Z=56$) are formed as a consequence of the bombardment of uranium ($Z=92$) with neutrons*". They have since established this fact beyond any shadow of doubt [82]. These results suggested that after neutron capture, a uranium nucleus splits up into two nuclei of medium atomic weight (e.g. barium and krypton), giving rise to a new type of disintegration process—*nuclear fission*.

After division, the high neutron-proton ratio of uranium tends to readjust itself by β-decay to the lower values of the lighter elements. The large neutron excess of some of the product nuclei is thus reduced to normal values by successive β-transformations which give rise to radioactive series. Since, moreover, it seems likely on general grounds that the exact ratio of the masses of the fission products may vary within somewhat wide limits, it can be readily appreciated that several β-radioactive isotopes of any given element may be produced and, further, that parallel β-radioactive series may arise. The fission process thus gives a natural explanation of the previous difficulty of the large number of isomers which were postulated to relate the transformations of the trans-uranic elements and of the parallel isomeric series.

It was suggested by Joliot that it is likely that the fission products are normally formed in a highly excited state, so that at first a small number

of neutrons will be evaporated. There is now good experimental evidence for several successive β-transformations and for neutron evaporation during fission. These results will, however, be discussed in more detail later.

Since the discovery of the new disintegration process, numerous investigations have been carried out, and it has been proved conclusively by several methods that the disruption of uranium and thorium nuclei into two lighter nuclei does take place when these substances are bombarded with neutrons.

It was at first sight difficult to understand how such a fission could occur, but it was soon suggested by Meitner and Frisch [83] that the process is readily explicable on the basis of the Bohr nuclear model. It was pointed out that, on account of the close packing of the constituent particles in a heavy nucleus and the consequent strong energy exchange between them, such a nucleus would be expected to behave in a collective way resembling the behaviour of a liquid drop. In consequence, when the heavy nucleus is sufficiently highly excited, it will break up into two lighter nuclei, just as a liquid drop will divide into two smaller drops if enough energy is given to it.

After the actual splitting, the product nuclei will repel each other, and it can be calculated, on the basis of the known nuclear radii and charge numbers, that the two fission products should thus gain a total kinetic energy of $\sim$200 Mev. This energy of course is actually available owing to the difference in packing fraction between uranium and product nuclei near the middle of the periodic table. In this way the fission process can be explained in an essentially classical way without the necessity of considering quantum mechanical effects, which would be extremely small on account of the large masses involved.

An estimate of the energy of the recoiling fragments can be made if it is remembered that the energy available as kinetic energy of the heavy particles will be reduced by 30–40 Mev. due to the emission of β- and γ-radiation. Such estimates show that the kinetic energy of the individual fission products is of the order 70 Mev., which is sufficient for some of them to pass through a thin layer of uranium and to have, moreover, a range of several cm. of air. They can thus be recorded by the bursts of ionization they produce in ionization chambers and may be deposited on surfaces placed near the irradiated uranium or thorium by the method of radioactive recoil.

The heavily-ionizing recoil fission products were first observed by Frisch [84], using a uranium-lined ionization chamber connected to a linear amplifier. The amplifier was connected to a thyratron which was biased so as to count only pulses corresponding to more than 5×10^5 ion pairs. No such bursts of ionization were observed until a neutron source consisting of 300 mgm. of radium mixed with beryllium was placed near the ionization chamber, when about fifteen particles per minute were recorded. In addition, it was found that surrounding the source with paraffin wax enhanced the effect by a factor of 2.

The maximum size of pulse observed was 2×10^6 ion pairs, corresponding to an energy loss of 70 Mev. (for the ions were formed on a path equivalent to 0·8 cm. of air or less). From this result, Frisch estimated that the particles concerned must have an atomic weight greater than 70. Thus conclusive

physical evidence of the breaking up of the uranium nucleus into two parts of comparable size was obtained, as indicated by the experiments of Hahn and Strassmann.

Experiments with thorium instead of uranium gave quite similar results, except that surrounding the neutron source with paraffin did not enhance, but slightly diminished, the effect. This showed that some of the fission processes in uranium are produced by slow as well as by fast neutrons, whereas the disintegration of thorium is produced only by fast neutrons.

These interesting results were almost immediately independently obtained by a number of other workers, including Roberts, Meyer and Hafstad [85], and Green and Alvarez [86]. They have since been confirmed on many occasions.

Recently Booth, Dunning and Slack [87] have investigated the energy distribution of the fragments resulting from uranium fissions under well-defined conditions. Their results indicate that the recoil nuclei are divided into at least two major energy groups, there being approximately the same number of particles in each group (though the possibility of more than two groups cannot be excluded). The maximum energies observed are approximately 100 Mev. and 72 Mev., suggesting that the ratio of the masses of the two nuclei formed in a single fission process is approximately 96/140. The difference between the 175 Mev. observed as kinetic energy of the particles and the 210 Mev. available in the disintegration can be accounted for by the excitation of the resulting fragments resulting in β, γ and neutron emission.

The tracks of the recoiling nuclei have also been photographed in Wilson chambers operating at reduced pressure by Joliot [88] and by Corson and Thornton [89], and have been recorded in photographic emulsions impregnated with uranium salts [90].

The fission products can also be collected as they recoil, by placing receiving surfaces near uranium or thorium which is being irradiated with neutrons.

Joliot [91] performed such an experiment by placing a neutron source thinly coated with uranium inside a concentric cylinder of bakelite. He showed that the bakelite did not become active under the action of neutrons in the absence of the uranium and, moreover, that it remained inactive in the presence of uranium and the absence of neutrons. In the presence of uranium and the source, however, the interior surface of the bakelite received a complex mixture of radioactive atoms which decayed in a manner analogous to that of the radio-elements formed in uranium.

By interposing thin cylindrical screens between the uranium and the bakelite, Joliot found that the range of the recoil atoms is of the order of 3 cm. of air. Experiments showed also an analogous effect in the case of thorium.

This method constitutes a simple physical method of separating the artificial radio-elements formed in uranium or thorium. It has been used by Meitner and Frisch [92], who collected the fission products on water and separated the active isotopes chemically. Platinum chloride was added to the solution and was precipitated chemically as sulphide. For comparison a sample of uranium was separated for " trans-uranic " activities by the same process.

The two decay curves were in perfect agreement with each other and with previous results obtained by Hahn, Meitner and Strassmann. Meitner and Frisch concluded from these experiments that *the "trans-uranic" activities originate by the cleavage of the uranium nucleus, and so must be ascribed to elements considerably lighter than uranium.*

Much of the more definite evidence for the production of elements of medium atomic weight by the fission process has been obtained by careful chemical analysis, whereby numerous radioactive bodies of diverse chemical properties have been separated from irradiated uranium and thorium. In this connection the work of Hahn and Strassmann on the radioactive isotopes of barium so produced is outstanding, for it demonstrated in a most convincing manner the occurrence of nuclear fission. It is worthy of note that this excellent work was carried out with neutron sources of low intensity.

However, some of the most conclusive physical evidence for the phenomenon was obtained by Abelson [93], and independently by Feather and Bretscher [94], for these investigators observed the emission of *iodine* Kα-radiation from the 72-hour "*trans-uranic*" activity, which was then shown by chemical tests to be a radioactive isotope of tellurium. (The iodine Kα-radiation proceeds from internal conversion of γ-radiation from the product iodine nucleus formed by the β-transformation.)

The experiments of Abelson [95], who has made an extensive study of the radioactivity of isotopes of antimony, tellurium and iodine produced by the fission of uranium, are also noteworthy. Abelson has carefully compared the radioactive tellurium and iodine isotopes formed from uranium by neutrons with those produced by the bombardment of tellurium with high-speed deuterons. By making detailed measurements of the half-lives of the active bodies and of the energies of the emitted radiations, and by using the technique of successive extractions, he has identified many of the fission products and has shown, in a striking fashion, that in several cases the same isotopes can be formed by both processes.

Other interesting experiments have been done by extracting noble gases from solutions of uranium and thorium salts [96], [97], [98]. In this way radioactive isotopes of krypton and xenon have been identified and their transformation products have been obtained as active deposits.

However, so many activities have now been identified that it is convenient to summarize the results in the form of the accompanying table 4. Many half-lives other than those given have been reported, but in most cases these have been obtained by the analysis of complex decay curves, due to a mixture of several radioactive isotopes. The work of Bjerge, Brostrøm and Koch [99] and of Frisch [100] has shown that in these cases it is not really possible to obtain reliable values by analysing the complex decay curve obtained, for the large number of periods tends to eliminate any individual features of the curve. Such results have, therefore, been omitted. Several of these bodies spontaneously transform into active products, forming short radioactive series. Appended are a number of such series, with details of the various transformation products where these are known.

1. $^{83}Br \rightarrow \beta \rightarrow {}^{83}Kr^{*}$ (2·5 hrs.) $\rightarrow \gamma \rightarrow {}^{83}Kr$ (113 min.)
2. $^{88}Br \rightarrow \beta \rightarrow {}^{88}Kr$ (?) $\rightarrow \beta \rightarrow {}^{88}Rb$ (2·5 hrs.) $\rightarrow \beta \rightarrow {}^{88}Sr$ (18 min.)
3. Sr (6 hrs.) $\rightarrow \beta \rightarrow$ Y $\rightarrow \beta \rightarrow$ Zr (3·5 hrs.) $\rightarrow \beta \rightarrow$ Nb (3·5 hrs (?))
4. Sb (5 min.) $\rightarrow \beta \rightarrow$ Te $\rightarrow \beta \rightarrow$ I (77 hrs.) $\rightarrow \beta \rightarrow$ Xe (2·4 hrs.)
5. Sb (10 min.) $\rightarrow \beta \rightarrow$ Te $\rightarrow \beta \rightarrow$ I (43 min.) $\rightarrow \beta \rightarrow$ Xe (54 min.)
6. Sb (10 min.) $\rightarrow \beta \rightarrow$ Te $\rightarrow \beta \rightarrow$ I (60 min.) $\rightarrow \beta \rightarrow$ Xe (22 hrs.)
7. $^{139}Xe \rightarrow \beta \rightarrow {}^{139}Cs$ (~0·5 min.) $\rightarrow \beta \rightarrow {}^{139}Ba$ (30 min.) $\rightarrow \beta \rightarrow {}^{139}La$ (87 min.)
8. Cs $\rightarrow \beta \rightarrow$ Ba (?) $\rightarrow \beta \rightarrow$ La (14 min.) $\rightarrow \beta \rightarrow$ Ce (2·5 hrs.)
9. Xe (~0·5 min.) $\rightarrow \beta \rightarrow$ Cs $\rightarrow \beta \rightarrow$ Ba (33 min.) $\rightarrow \beta \rightarrow$ La (250 hrs.) $\rightarrow \beta \rightarrow$ Ce (3 hrs.)

Table 4. *Fission products from uranium and thorium* *

Element (atomic number)	Mass number	Half-life	Emitted radiations	Method of formation	Classification (Livingood and Seaborg)
Selenium	—	Several hours	—	$Th+{}^{1}n$	D
(34)		Several days	—	$Th+{}^{1}n$	D
Bromine	83	2·5 hrs.	$\bar{e}$	$Th+{}^{1}n$	A
(35)		40 min.	—	$U+{}^{1}n$	D
		22 hrs.	—	$Th+{}^{1}n$	F
Krypton	88	3·5 hrs.	$\bar{e}$	$Th+{}^{1}n$	A
(36)	83	113 min.	i.c.$\bar{e}$	$Th+{}^{1}n$ β-decay of ^{83}Br	A
Rubidium	88	18 min.	$\bar{e}$	$U+{}^{1}n$; $Th+{}^{1}n$ β-decay of ^{88}Kr	A
(37)					
Strontium	89	55 days	$\bar{e}$	$U+{}^{1}n$	A
(38)	>90	6 hrs.	$\bar{e}$	$U+{}^{1}n$	D
		7 min.	—	$U+{}^{1}n$	D
Yttrium	>90	3·5 hrs.	$\bar{e}$	$U+{}^{1}n$; $Th+{}^{1}n$ (?) β-decay Sr 6 hrs.	D
(39)					
Molybdenum	99 or 101	66 hrs.	$\bar{e}$	$U+{}^{1}n$	C
(42)					
Antimony	127	82 hrs.	$\bar{e}$	$U+{}^{1}n$	A
(51)	129	4·6 hrs.	$\bar{e}$	$U+{}^{1}n$	A
	>131	<15 min.	$\bar{e}$	$U+{}^{1}n$	A
	>131	5 min.	$\bar{e}$	$U+{}^{1}n$	D
		40 min.	$\bar{e}$	$U+{}^{1}n$	D

* There are other activities not yet identified.

Table 4 (*cont.*)

Element (atomic number)	Mass number	Half-life	Emitted radiations	Method of formation	Classification (Livingood and Seaborg)
Tellurium (52)	127	10 hrs.	$\bar{e}$	$U+{}^1n$	A
	129	70 min.	$\bar{e}$	$U+{}^1n$	A
				β-decay ^{129}Sb	
	131	30 hrs. }	$\bar{e}$	$U+{}^1n$	A
		25 min. }			
	>131	40 min.	$\bar{e}$	$U+{}^1n$	D
	>131	1 hr.	$\bar{e}$	$U+{}^1n$	D
	>131	72 hrs.	$\bar{e}$	$U+{}^1n$	D
				β-decay Sb 15 min.	
Iodine (53)	131	8·0 days	$\bar{e}$	$U+{}^1n$	A
				β-decay ^{131}Te	
	>131	2·5 hrs.	$\bar{e}$	$U+{}^1n$	
				β-decay Te 72 hrs.	
	>131	54 min.	$\bar{e}$	$U+{}^1n$	
				β-decay Te 40 min.	
	>131	22 hrs.	$\bar{e}$	$U+{}^1n$	
				β-decay Te 1 hr.	
Xenon (54)	>131	5·5 days	$\bar{e}$	$Th+{}^1n$	
				β-decay I 22 hrs.	
	139	5 min.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	
	>139	~0·5 min.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	
		~1 day (?)	—	$Th+{}^1n$	
Caesium (55)	139	8 min.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	A
				β-decay ^{139}Xe	
	>139	35 min.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	D
				β-decay Xe ~0·5 min.	
Barium (56)	139	86 min.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	A
				β-decay ^{139}Cs	
		3 min.	—	$U+{}^1n$; $Th+{}^1n$	F
	? 140	250 hrs.	$\bar{e}$	$U+{}^1n$	D
	>140	14 min.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	D
	>140	21 min.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	E
Lanthanum (57)	140	31 hrs.	$\bar{e}$	$U+{}^1n$	A
				β-decay ^{140}Ba (?)	
	?>140	3·5 hrs.	$\bar{e}$	$U+{}^1n$; $Th+{}^1n$	D
				β-decay Ba 14 min.	
	>140	30 min.	$\bar{e}$	$U+{}^1n$	
				β-decay Ba <1 min.	
Rare earth (>57)	—	2·5 days	$\bar{e}$	$U+{}^1n$	F

The question now arises whether there is any evidence for the formation of elements of higher atomic number than uranium by irradiating uranium with neutrons, or whether all the trans-uranic activities are to be ascribed to lighter nuclei produced by fission. (As already mentioned, the experiments of Meitner and Frisch suggest that all the trans-uranic bodies are in reality of medium atomic weight.)

In this connection some results obtained by McMillan [101] and Segrè [102] are significant. McMillan mounted a very finely divided uranium source on filter paper and collected the recoil products on very thin paper. He observed that some of the radioactivities induced by neutrons remained in the sample and that the decay curve showed quite striking differences from that of the fission products. The uranium curve showed a strong component of half-life about 25 minutes and also a long period of 2·3 days.

The activity of half-life 23 minutes has been identified by Hahn, Meitner and Strassmann as a uranium isotope, and it is produced by a resonance neutron-capture process. It is undoubtedly due to ^{239}U, and following its decay there must be produced an isotope of eka-rhenium ($Z=93$).

Since the trans-uranic activities would certainly be found in the non-recoiling fraction, a detailed investigation of the two activities thus observed by McMillan was made by Segrè. It seemed possible, at first sight, that the body of 2-day period might be the transformation product of that with the 23-minute period, but direct observations of the decay of the chemically separated uranium excluded this possibility.

Careful chemical analysis showed that the body of half-life 2·3 days is actually a rare earth, probably considerably heavier than lanthanum.

These experiments have therefore shown conclusively that no β-radioactive trans-uranic activities have yet been observed.

However, the ^{239}U decays into a substance of atomic number 93. Accordingly Segrè looked carefully for α-ray emission, but no such short-lived activity was detected. Moreover, the eka-rhenium separated from uranium which had been activated with neutrons from the Berkeley cyclotron for two years was completely inactive, from which it *must be concluded that ^{239}U decays into a very long-lived isotope of eka-rhenium and that trans-uranic elements have not yet been observed.*

More recently, Hahn [103] also inclines to this view, for he has proved that his "eka-iridium" of half-life 2·5 hours is in reality an isotope of iodine and that the 66-hour "eka-platinum" consists of two active bodies with the chemical properties of tellurium and molybdenum respectively.

The production of a radioactive isotope of thorium (analogous to the 23-minute ^{239}U) by simple neutron capture has been observed by Meitner, Hahn and Strassmann [104]. The capture processes for the formation of these bodies are essentially resonance phenomena. Thus in uranium, for neutron energies of the order of 25 volts, the capture cross-section is at least 30 times larger than that for thermal neutrons. Since the cross-section in this region amounts to 10^{-21} cm^2 it is obvious that this process must occur with the abundant uranium isotope of mass 238. In neither uranium nor thorium is the

resonance capture accompanied by any large increase in the cross-section for the fission processes ($\sim 2 \times 10^{-24}$ cm^2), showing that the compound state of ^{239}U formed by neutron capture must have a life-time 10^6 times longer than the time the nucleus would take to divide itself. Thus the probability of radiation by these compound nuclei is considerably larger than the fission probability, and their normal states are therefore essentially " stable ".

It is then interesting to note that with fast neutrons, fission cross-sections of the same order of magnitude were found both for uranium and thorium, but with thermal neutrons a large increase of the fission cross-section was observed for the former but not for the latter. Moreover, it was shown by Anderson *et al.*[105] that the behaviour of the slow neutrons causing the fission process is similar to that of those causing the boron disintegration, suggesting that the efficiency of the slow neutrons for the fission process follows a $1/v$ law. Bohr[106] has explained the apparent contradiction between this result and the observed sharp capture resonance at 25 volts by suggesting that the fast neutron fissions and the resonance capture process occur with ^{238}U, whereas the fissions due to thermal neutrons arise by the disintegration of the highly excited ^{236}U formed by the capture of neutrons by the rare isotope actino-uranium ^{235}U.

The time involved in the fission process has been shown to be less than 3×10^{-3} sec. by Green and Alvarez[107] using a modulated beam of neutrons. Feather[108], however, has performed experiments from which he has deduced a very much smaller limiting time than this. This result was obtained by studying the spatial asymmetry of the disintegration products of uranium bombarded by fast neutrons. The asymmetry observed shows that after capture of the neutron, fission usually occurs before the original momentum of the system has been lost. The time involved in the process can then be estimated from the known data on the ranges of heavy radioactive recoil atoms. In this way Feather deduced that the mean life of the intermediate nucleus could not be longer than about 5×10^{-13} sec.

As the fission products are normally formed in a very highly excited state it was suggested by Joliot as likely that at first the large neutron excess will be partially reduced by the direct liberation of neutrons occurring at the instant of fission or as an evaporation from the resulting newly formed nuclei, a process which would appear to be the more probable on account of the large neutron excess of the fragments. The first evidence for this phenomenon was obtained by von Halban, jun., Kowarski and Joliot[109], who showed that secondary neutrons are actually produced in the interaction between uranium nuclei and neutrons and, moreover, that the number of neutrons thus produced is sufficient at least to compensate for the absorption of thermal neutrons in uranium. It was observed that the cross-section for the production of these neutrons is greater than the fission cross-section, suggesting that several neutrons are probably emitted during each fission, a condition which is necessary for the possibilities of chain reactions.

It was shown, moreover, by Dodé, von Halban, jun., and Kowarski[110] that the neutrons thus produced are qualitatively distinct from the incident

neutrons. In these experiments a source of photo-neutrons consisting of 300 mgm. of radium inside a beryllium block $5 \times 6 \times 2$ cm. was surrounded with uranyl nitrate (10 gm./cm^2). The whole was placed in a vessel containing 8 litres of carbon bisulphide in which was dissolved 200 mgm. of phosphorus. After 6 days' irradiation the phosphorus was isolated and its activity measured with a Geiger-Müller counter. The activity (35 counts/min.) was due to ^{32}P.

The reaction ^{32}S (n, p) ^{32}P is endothermic by 0·9 Mev., and for it to be produced with a finite probability, neutrons of energy >2 Mev. are necessary. The photo-neutrons from the source have much too low an energy to excite this reaction and, moreover, the more energetic neutrons from the reactions between the α-particles from the radium and the glass container caused only 1/6 the total effect. Thus since the detector was insensible to the primary neutrons, this experiment showed in a direct manner the production in the uranium of " fast " neutrons by photo-neutrons.

von Halban, jun., Joliot and Kowarski [111] further concluded that the number of neutrons emitted per fission is $3 \cdot 5 \pm 0 \cdot 7$.

Confirmatory evidence has been obtained by Anderson, Fermi and Hanstein [112], Szilard and Zinn [113], Michiels, Parry and Thomson [114] and Rotblat [115]. While the estimates of the number of neutrons emitted are not too concordant, it is conceivable that more neutrons are produced by the fast neutron fissions as investigated by Rotblat than by those caused by thermal neutrons, for the neutron excess in the ^{238}U$+n$ reactions must be greater than in similar processes with ^{235}U$+n$. It seems likely, however, that more than one neutron is emitted per fission, the probable number lying between 1 and 3.

The experiments of Joliot *et al.* [116] have shown that neutrons of energy >2 Mev. are emitted during the fission of uranium, but these do not show whether neutrons of notably greater energy are present in the secondary radiation. A search was therefore made for neutrons of higher energy. Evidence was obtained with an oxygen-filled ionization chamber of recoil nuclei with energies as high as 2·5 Mev., indicating that neutrons with energy ~11 Mev. are liberated from uranium when irradiated with thermal neutrons.

The high energy of these fast neutrons shows that the nuclei from which they are ejected must be in a highly excited state at the moment of formation, and that with the ejection of the fast neutron a not negligible fraction of the fission energy is released.

The *delayed emission of neutrons* from uranium and thorium which had been activated with neutrons was first observed by Roberts, Meyer and Wang [117]. Using a boron-lined ionization chamber as recording apparatus, neutrons were observed several minutes after the activation had ceased. The decay period of the neutron emission was found to be $12 \cdot 5 \pm 3$ sec. Hard γ-rays with the same half-life were also observed, and it was at first suspected that the neutrons might be of secondary origin, being produced by photo-disintegration. This seemed very unlikely, however, in view of the very large cross-section needed to account for the observed number of neutrons.

Careful control experiments by Roberts, Hafstad, Meyer and Wang [118] proved, however, that these neutrons are in fact directly emitted by some of the

fission products. Cloud-chamber observations of the recoil protons produced by the delayed neutrons showed that the energy of the neutrons is less than 1·0 Mev., and is probably closer to 0·5 Mev.

Szilard and Zinn [119] have looked for the delayed emission of neutrons of much higher energy, but have concluded that the number must be much less than the number of fast neutrons emitted at the instant of fission.

In addition, it has been observed that the delayed neutrons (and all the γ-ray periods) are produced both by thermal and high-energy neutrons under the same conditions as the uranium fission. The cross-section for the delayed emission process has been estimated to be 4×10^{-26} cm^2.

Booth, Dunning and Slack [119] have extended the observations of Roberts *et al.* and have shown that the effect is primarily produced by slow neutrons which follow approximately the $1/v$ absorption law in boron. The decay curve of delayed neutron emission is best analyzed into two components of half-lives 45 sec. and 10–15 sec. respectively. β-rays (energy >3 Mev.) and hard γ-rays of the same half-life have also been detected. These observations seem to indicate that some of the fission fragments emit β-rays of these periods, and the product nuclei are then sufficiently excited to emit high energy γ-ray quanta or, less frequently, "delayed" neutrons. Booth, Dunning and Slack estimate that the ratio of the initial number of delayed neutrons per second to the total number of fissions per second is approximately 1 to 60, showing that the delayed neutrons are much less numerous than the fast neutrons evaporated immediately after fission occurs.

An interesting possibility which has to be considered, and which has been the subject of considerable conjecture and discussion since the discovery of the fission process and the observation that more than one neutron is produced per fission, is that a cumulative process of exothermic disintegration or chain reaction may occur. It is quite clear that unless other processes than those which result in fission can take place, the latter process could "build up" in large quantities of uranium. As yet, however, direct experimental evidence for this process is, perhaps fortunately, not available, and the possible catastrophic explosions or self-perpetuating reaction chains have not been observed.

It must be remembered, however, as has been pointed out by Feather [120], that even in pure uranium the non-fission capture process occurs, and, moreover, the unlimited generation of energy in the solid material would ultimately increase the energy of the thermal neutrons until their efficiency as agents of fission was greatly reduced. Several attempts have been made to predict what might occur if sufficient quantities of uranium were irradiated with neutrons, but the assumptions on which they have been based have, in general, been so idealized that one can scarcely have much confidence in the predictions.

In conclusion I wish to express my thanks to Professor J. Chadwick, F.R.S., who has been kind enough to read the manuscript of this article. The helpful advice he has given me during its preparation I acknowledge gratefully. I would like also to thank Professor Luis W. Alvarez and Dr. Emilio Segrè of the University of California for assisting me in a similar manner.

I am grateful to Dr. Neary for sending me, before publication, details of his work on the spectrum of RaE.

REFERENCES

§ 1. *Introduction*

(1) Curie and Joliot. *C.R. Acad. Sci., Paris*, **198**, 254 (1934).
(2) Livingood and Seaborg. *Rev. Mod. Phys.* (Jan. 1940).

§ 2. *The range : energy relation for β-rays*

(3) Widdowson and Champion. *Proc. Phys. Soc.* **50**, 185 (1938).
(4) Feather. *Proc. Camb. Phil. Soc.* **34**, 599 (1938).

§ 3. *The continuous β-ray spectrum*

(5) Fermi. *Z. Phys.* **88**, 160 (1934).
(6) Uhlenbeck and Konopinski. *Phys. Rev.* **48**, 7 (1935).
(7) Paxton. *Phys. Rev.* **51**, 170 (1937).
(8) Flammersfeld. *Z. Phys.* **112**, 727 (1939).
(9) Neary. Thesis, Cambridge (1938).
(10) Martin and Townsend. *Proc. Roy. Soc.* A, **170**, 190 (1939).
(11) Lawson. *Phys. Rev.* **56**, 125 (1939).
(12) Tyler. *Phys. Rev.* **56**, 131 (1939).
(13) Langer and Whittaker. *Phys. Rev.* **51**, 713 (1937).
(14) Lyman. *Phys. Rev.* **51**, 1 (1937).
(15) Lyman. *Phys. Rev.* **55**, 234 (1939).
(16) Bethe, Hoyle and Peierls. *Nature, Lond.*, **143**, 200 (1939).
(17) Alichanian and Berestezky. *Phys. Rev.* **55**, 978 (1939).

§ 4. *Nuclear K-electron capture*

(18) Møller. *Phys. Z. Sowjet*, **11**, 9 (1937).
(19) Alvarez. *Phys. Rev.* **52**, 134 (1937).
(20) Alvarez. *Phys. Rev.* **54**, 486 (1938).
(21) Walke, Williams and Evans. *Proc. Roy. Soc.* A, **171**, 360 (1939).
(22) Roberts, Heydenburg and Locher. *Phys. Rev.* **53**, 1016 (1938).
(23) Thompson and Walke. Unpublished.
(24) Walke, Thompson and Holt. Unpublished.
(25) Hemmendinger. *Phys. Rev.* **55**, 604 (1939).
(26) Livingood and Seaborg. *Phys. Rev.* **54**, 391 (1938).
(27) Livingood and Seaborg. *Phys. Rev.* (Feb. 1940).
(28) Livingood and Seaborg. *Phys. Rev.* **55**, 457 (1939).
(29) Alvarez. *Phys. Rev.* **54**, 486 (1938).
(30) Cacciapuoti. *Phys. Rev.* **55**, 110 (1939).
(31) Richardson. *Phys. Rev.* **55**, 609 (1939).
(32) Ridenour, Delsasso, White and Sherr. *Phys. Rev.* **53**, 770 (1938
(33) Seaborg, Livingood and Kennedy. *Phys. Rev.* **55**, 794 (1939).
(34) Richardson. *Phys. Rev.* **55**, 609 (1939).
(35) Oldenberg. *Phys. Rev.* **53**, 35 (1938).
(36) Alvarez. *Phys. Rev.* **54**, 486 (1938).

§ 5. *Nuclear isomerism*

(37) Soddy. *Nature, Lond.*, **99**, 414 and **433** (1917).
(38) Hahn. *Ber. dtsch. Chem. Ges.* B, **54**, 1131 (1921).
(39) Feather and Bretscher. *Proc. Roy. Soc.* A, **165**, 530 (1938).

(40) Weizsäcker. *Naturwissenschaften*, **28**, 813 (1936).
(41) Pontecorvo. *Reports of Congress at the Palais des Découverts, Paris*, 1938.
(42) Hebb and Uhlenbeck. *Physica*, v, **7**, 605 (1938).
(43) Pontecorvo. *Phys. Rev.* **54**, 542 (1938).
(44) Segrè, Halford and Seaborg. *Phys. Rev.* **55**, 321 (1939).
(45) Szilard and Chalmers. *Nature, Lond.*, **134**, 462 (1934).
(46) De Vault and Libby. *Phys. Rev.* **35**, 322 (1939).
(47) Le Roux, Lu and Sugden. *Nature, Lond.*, **143**, 518 (1939).
(48) Valley and McCreary. *Phys. Rev.* **55**, 666 (1939).
(49) Siday. *Nature, Lond.*, **143**, 681 (1939).
(50) Abelson. *Phys. Rev.* **55**, 424 (1939).
(51) Roussinow and Yusephovich. *Phys. Rev.* **55**, 979 (1939).
(52) Seaborg, Livingood and Kennedy. *Phys. Rev.* **55**, 794 (1939).
(53) Feather and Dunworth. *Proc. Roy. Soc.* A, **168**, 566 (1938).

List of isomers

A. Burcham, Goldhaber and Hill. *Nature, Lond.*, **141**, 510 (1938).
Walke. *Phys. Rev.* **52**, 400 (1937).
Dubridge, Barnes, Buck and Strain. *Phys. Rev.* **53**, 447 (1938).
Gentner. *Naturwissenschaften*, **26**, 109 (1938).
Walke. *Phys. Rev.* **52**, 777 (1937).
B. Walke. *Phys Rev* **52**, 777 (1937).
Amaldi *et al.* *Proc. Roy. Soc.* A, **149**, 522 (1935).
Walke, Williams and Evans. *Proc. Roy. Soc.* A, **171**, 360 (1939).
C. Darling, Curtis and Cork. *Phys. Rev.* **51**, 1011 (1937).
Livingood and Seaborg. *Phys. Rev.* **54**, 391 (1938).
Hemmendinger. *Phys. Rev.* **55**, 604 (1939).
D. Livingood and Seaborg. *Phys. Rev.* **55**, 457 (1939).
Heyn. *Physica*, **4**, 1224 (1937).
Thornton. *Phys. Rev.* **53**, 326 (1938).
E. Snell. *Phys. Rev.* **50**, 1007 (1937).
Segrè, Halford and Seaborg. *Phys. Rev.* **55**, 32 (1939).
Valley and McGreary. *Phys. Rev.* **55**, 666 (1939).
Alichanian, Alichanow and Dzelepow. *Phys. Z. Sowjet*, **10**, 78 (1936).
Bothe and Gentner. *Naturwissenschaften*, **31**, 517 (1938).
Pool, Cork and Thornton. *Phys. Rev.* **52**, 239 (1937).
F. Pontecorvo. *Phys. Rev.* **54**, 542 (1938).
Crittenden, jun., and Bacher. *Phys. Rev.* **54**, 862 (1938).
Brieben. *Naturwissenschaften*, **138**, 1095 (1930).
G. Pool. *Phys. Rev.* **53**, 116 (1938).
Bothe and Gentner. *Z. Phys.* **112**, 45 (1939).
Pool, Cork and Thornton. *Phys. Rev.* **52**, 239 (1937).
H. Amaldi *et al.* *Proc. Roy. Soc.* A, **149**, 522 (1935).
Gaerttner, Turin and Crane. *Phys. Rev.* **49**, 793 (1936).
Mitchell and Langer. *Phys. Rev.* **53**, 505 (1938).
Lawson and Cock. *Phys. Rev.* **52**, 531 (1937).
Curtis and Richardson. *Phys. Rev.* **53**, 942 (1938).
Barnes. *Phys. Rev.* **55**, 241 (1939).
I. Seaborg, Livingood and Kennedy. *Phys. Rev.* **55**, 794 (1939).
Abelson. *Phys. Rev.* **55**, 876 (1939).
Tape and Cork. *Phys. Rev.* **53**, 676 (1938).
J. Seaborg, Livingood and Kennedy. *Phys. Rev.* **55**, 794 (1939).
Bothe and Gentner. *Naturwissenschaften*, **25**, 90 (1937).
Heyn. *Nature, Lond.*, **139**, 842 (1937).
Abelson. *Phys. Rev.* **55**, 876 (1939).

K. Seaborg, Livingood and Kennedy. *Phys. Rev.* **55**, 794 (1939).
Abelson. *Phys. Rev.* **55**, 876 (1939).
L. Amaldi *et al.* *Proc. Roy. Soc.* A, **149**, 522 (1935).
Alexeeva. *C.R. Acad. Sci., U.R.S.S.*, **8**, 553 (1938).
M. Hevesy and Levi. *Nature, Lond.*, **137**, 185 (1936).
Marsh and Sugden. *Nature, Lond.*, **136**, 102 (1935).
N. Bothe and Gentner. *Naturwissenschaften*, **25**, 90 (1937).
Oldenberg. *Phys. Rev.* **53**, 35 (1938).
O. McMillan, Kamen and Ruben. *Phys. Rev.* **52**, 375 (1937).
P. McMillan, Kamen and Ruben. *Phys. Rev.* **52**, 375 (1937).
Q. Hahn. *Ber. dtsch. Chem. Ges.* B, **54**, 1131 (1921).
Feather and Bretscher. *Proc. Roy. Soc.* A, **165**, 530 (1938).

§ 6. *Induced radioactivity by nuclear excitation*

(54) Dodé and Pontecorvo. *C.R. Acad. Sci., Paris*, **207**, 287 (1938).
(55) Goldhaber, Hill and Szilard. *Phys. Rev.* **55**, 47 (1939).
(56) Cork and Lawson. *Phys. Rev.* **55**, 1136 (1939).
(57) Barnes and Aradine. *Phys. Rev.* **55**, 50 (1939).
(58) Lark-Horovitz, Risser and Smith. *Phys. Rev.* **55**, 878 (1939).
(59) Pontecorvo and Lazard. *C.R. Acad. Sci., Paris*, **208**, 99 (1939).
(60) Collins, Waldman, Stubblefield and Goldhaber. *Phys. Rev.* **55**, 507 (1939).
(61) Barnes. *Bull. Amer. Phys. Soc.*, Stanford Meeting, Abst. 55 (1939).
(62) Kalbfell. *Phys. Rev.* **54**, 543 (1938).
(63) Seaborg and Segrè. *Phys. Rev.* **55**, 808 (1939).
(64) Langsdorf and Segrè. *Phys. Rev.* (1940).

§ 7. *Element 43*

(66) Perrier and Segrè. *J. Chem. Phys.* **5**, 712 (1937).
(67) Geilmann and Weibke. *Z. anorg. Chem.* **199**, 120 (1931).
(68) Perrier and Segrè. *J. Chem. Phys.* **7**, 155 (1939).
(69) Cacciapuoti and Segrè. *Phys. Rev.* **52**, 1252 (1937).
(70) Seaborg and Segrè. *Phys. Rev.* **55**, 808 (1939).
(71) Walke. *Phil. Mag.* **19**, 878 (1935).
(72) Jensen. *Naturwissenschaften*, **26**, 381 (1938).
(73) Abelson. Unpublished.

§ 8. *Induced radioactivity by hard γ-rays (nuclear photo-disintegration)*

(74) Bothe and Gentner. *Naturwissenschaften*, **25**, 90, 126, 191 and 284 (1937).
(75) Chang, Goldhaber and Sagane. *Nature, Lond.*, **139**, 962 (1937).
(76) Bothe and Gentner. *Z. Phys.* **112**, 45 (1939).
(77) Fowler, Gaerttner and Lauritsen. *Phys. Rev.* **53**, 628 (1938).

§ 9. *Induced radioactivity by the fission of uranium and thorium nuclei*

(78) Fermi *et al.* *Proc. Roy. Soc.* A, **146**, 483 (1934).
(79) Meitner, Hahn and Strassmann. *Z. Phys.* **106**, 249 (1937).
(80) Curie and Savitch. *J. Phys. Radium*, **9**, 355 (1938).
(81) Hahn and Strassmann. *Naturwissenschaften*, **26**, 756 (1938).
(82) Hahn and Strassmann. *Naturwissenschaften*, **27**. 2 (1939).
(83) Meitner and Frisch. *Nature, Lond.*, **183**, 239 (1939).
(84) Frisch. *Nature, Lond.*, **143**, 276 (1939).
(85) Roberts, Meyer and Hafstad. *Phys. Rev.* **55**, 416 (1939).
(86) Green and Alvarez. *Phys. Rev.* **55**, 617 (1939).
(87) Booth, Dunning and Slack. *Phys. Rev.* **55**, 981 (1939).

(88) JOLIOT. *C.R. Acad. Sci., Paris*, **208**, 647 (1939).
(89) CORSON and THORNTON. *Phys. Rev.* **55**, 509 (1939).
(90) MYSSOWSKY and JDANOFF. *Nature, Lond.*, **143**, 794 (1939).
(91) JOLIOT. *C.R. Acad. Sci., Paris*, **208**, 341 (1939).
(92) MEITNER and FRISCH. *Nature, Lond.*, **143**, 471 (1939).
(93) ABELSON. *Phys. Rev.* **55**, 418 (1939).
(94) FEATHER and BRETSCHER. *Nature, Lond.*, **143**, 516 (1939).
(95) ABELSON. *Phys. Rev.* **56**, 1 (1939).
(96) HEYN, ATEN, jun. and BAKKER. *Nature, Lond.*, **143**, 516 (1939).
(97) ATEN, jun., BAKKER and HEYN. *Nature, Lond.*, **143**, 679 (1939).
(98) LANGSDORF. *Phys. Rev.* **56**, 205 (1939).
(99) BJERGE, BRØSTROM and KOCH. *Nature, Lond.*, **143**, 794 (1939).
(100) FRISCH. *Nature, Lond.*, **143**, 859 (1939).
(101) MCMILLAN. *Phys. Rev.* **55**, 510 (1939).
(102) SEGRÈ. *Phys. Rev.* **55**, 1104 (1939).
(103) HAHN and STRASSMANN. *Naturwissenschaften*, **27**, 451 (1939).
(104) MEITNER, HAHN and STRASSMANN. *Z. Phys.* **109**, 538 (1939).
(105) ANDERSON *et al.* *Phys. Rev.* **55**, 511 (1939).
(106) BOHR. *Phys. Rev.* **55**, 418 (1939).
(107) GREEN and ALVAREZ. *Phys. Rev.* **143**, 597 (1939).
(108) FEATHER. *Nature, Lond.*, **143**, 597 (1939).
(109) VON HALBAN, jun., KOWARSKI and JOLIOT. *Nature, Lond.*, **143**, 470 (1939).
(110) DODÉ, VON HALBAN, jun., and KOWARSKI. *C.R. Acad. Sci., Paris*, **208**, 995 (1939).
(111) VON HALBAN, jun., KOWARSKI and JOLIOT. *Nature, Lond.*, **143**, 680 (1939).
(112) ANDERSON, FERMI and HANSTEIN. *Phys. Rev.* **55**, 797 (1939).
(113) SZILARD and ZINN. *Phys. Rev.* **55**, 799 (1939).
(114) MICHIELS, PARRY and THOMSON. *Nature, Lond.*, **143**, 760 (1939).
(115) ROTBLAT. *Nature, Lond.*, **143**, 852 (1939).
(116) VON HALBAN, jun., JOLIOT and KOWARSKI. *Nature, Lond.*, **143**, 939 (1939).
(117) ROBERTS, MEYER and WANG. *Phys. Rev.* **55**, 510 (1939).
(118) ROBERTS, HAFSTAD, MEYER and WANG. *Phys. Rev.* **55**, 664 (1939).
(119) BOOTH, DUNNING and SLACK. *Phys. Rev.* **55**, 876 (1939).
(120) FEATHER. *Nature, Lond.*, **143**, 877 (1939).

SEPARATION OF ISOTOPES

By HAROLD C. UREY
Columbia University

§1. INTRODUCTION

When Soddy [1] first suggested the existence of isotopes, the experience of chemists supplied abundant evidence that the chemical properties of isotopes of an element were identical. Had this not been the case, differences in the combining weights of the elements should have been observed. Theories at the time suggested that unidirectional evaporation and diffusion should result in a separation of the isotopes, since the velocity of both processes should be inversely proportional to the square roots of the molecular weights. Such methods were first used by Aston [2], Harkins [3] and Brönsted and Hevesy [4] for producing slight changes in the relative abundances of isotopes. Simple calculation will show that the fractionation factors for such processes are so slight that simple diffusion or simple unidirectional evaporation could not be expected to produce any greatly increased concentration of isotopes, and these early experiments only prove that the atomic weights of the elements could be changed by such physical processes. After some 20 years, we now know that simple distillation of water would produce changes in the density of water due to the fractionation of even the rare isotopes of hydrogen and oxygen. Of course had this effect been observed, it would have been so small (only a few parts per million in the density) that impurities of one kind or another would have been suspected as the cause. That such variations of atomic weights were not observed was due to a combination of what might be called accidental circumstances. The heavy isotopes of hydrogen, carbon, nitrogen or oxygen are so rare that only slight changes in physical properties could be expected, and hence were overlooked. Elements of higher atomic weight have but slight differences in their physical and chemical properties, and thus the isotopes of elements such as chlorine, which are of comparable abundances, will not have sufficient differences in physical and chemical properties to cause observable variations in atomic weights. Lithium and boron are two elements of low atomic weights having isotopes of comparable abundance, but also having physical properties which make an exact determination of atomic weights very difficult, and hence any fractionation would not be observed in such cases. Helium, beryllium and fluorine are pure isotopic elements (or have only very rare isotopic species), and no separation would have been observed in these cases.

During the last ten years, very considerable success has been attained in the separation of isotopes, and in fact very many of the isotopes of elements are being separated in sufficient quantities for chemical and physical research,

and it appears that in the near future isotopes of any element may be secured, if for any reason we need them for scientific research. In this review no attempt is made to refer to all of the methods which might be used for the separation of isotopes. It is limited to those processes which either have been used for a substantial concentration of isotopes or which appear to the writer as probable methods for use in the future.

No effort has been made in the *Report* to use consistent symbols throughout, but rather an attempt is made to use symbols in each section which have been used by research workers in that particular field, so that the reader of this article may easily refer to the original literature without the necessity of changing from one set of symbols to another.

§ 2. METHODS OF SEPARATION OF ISOTOPES

Chemical methods for the separation of isotopes may be classified into three groups, (1) electrolysis, (2) distillation and (3) chemical exchange reaction methods. The slight differences in chemical properties, which make these methods possible, decrease rapidly with increasing temperature and increasing atomic weights. These methods have been most useful up to the present in concentrating isotopes of the lighter elements, hydrogen, lithium, carbon, nitrogen and oxygen, but cannot be expected to be useful in other cases. (Sulphur may be an exception[5].)

1. *Electrolytic methods*

Almost immediately after the discovery of deuterium, Washburn began experiments aimed at the concentration of deuterium by electrolytic means, and Washburn and Urey [6] first reported that the water removed from commercial electrolytic cells used for the production of hydrogen and oxygen contained increased concentrations of deuterium. In the industry, cells (generally using nickel electrodes and potassium hydroxide solutions as electrolytes) are operated for long periods of time for the production of hydrogen and oxygen without changing the electrolyte in the cells. Water is added daily to these cells to make up for the water electrolyzed. Under these conditions an increased concentration can be secured, which is determined by the simple process fractionation factor, defined as the ratio of isotopes in the gas escaping from such a cell divided by the ratio of isotopes in the water from which it is produced.

Lewis and MacDonald [7] were the first to use this method for the production of nearly pure deuterium. The method consists in electrolyzing a large volume of water down to a small residue. The Rayleigh distillation formula,

$$\left(\frac{1-N}{1-N_0}\right)^{\left(\frac{1}{\alpha-1}\right)}\left(\frac{N}{N_0}\right)^{\left(\frac{\alpha}{\alpha-1}\right)}=\frac{W_0}{W}, \qquad \ldots.(1)$$

applies to this process, where α is the fractionation factor, and N and N_0 are the mole fractions of deuterium in the water when the volumes of water are W and W_0 respectively. A very usual value of the fractionation factor α is 6.

Following this early work a number of plants have been described for the production of deuterium [8]. All use essentially the same process. A dilute solution of an electrolyte (sodium or potassium hydroxide, sulphuric acid, barium hydroxide, or even neutral salts, but usually potassium hydroxide) is electrolyzed until it becomes too concentrated. The water is then removed, a dilute solution prepared and the process repeated. When the original volume is reduced by a factor of 9000 the residue is about 90 per cent D_2O, and when by 25,000 the residue is 99·9 per cent D_2O. It is necessary to recombine the gases evolved in order to conserve deuterium. This is done by burning the gases from a nozzle through which the gases flow at high velocities, thus preventing back explosions. Thirty or forty thousand ampere-hours are required to produce about one gram of heavy water. It should be noted here, as well as in the case of other isotopes and other methods of separation, that most of the expenditure of effort and energy is required to increase the concentration by a small factor, say 2 or 3. Thus in the case of a rare isotope, if a supply containing n times the natural concentration can be found or produced inexpensively, one can produce n times as much product with the same energy and effort. The electrolyte of industrial cells has served as such a crude material with increased concentration of deuterium.

Many studies have been made on the electrolytic separation of hydrogen and deuterium. These studies have often been in disagreement and in fact hardly ever in quantitative agreement. The earlier work was done with the primary objective of finding a good method for separating the isotopes, but more recently several researches dealing primarily with the theory of the separation have been published. The problem is closely related to the overvoltage of hydrogen, and no attempt can be made here to review this much-discussed subject.

Considerable variation in fractionation factors has been reported. Oddie [9] showed that the fractionation factor increased with current density. Topley and Eyring [10], Bell and Wolfenden [11] and Collie [12] reported fractionation factors varying from 3·4 to 7·6. Brown and Daggett [13] secured a value for α with nickel electrodes of 8·6 when uncorrected for water vapor carried over with the exit gases, and of 12 when this correction was made. This was the average for many large cells and should be a reliable value. Eucken and Bratzler [14] secured a factor of 17 by giving the cathode a previous anodic polarization. Washburn, Smith and Frandsen [15] found a factor of 3 when low current densities were used and the electrodes were platinized platinum. This is very nearly the equilibrium fractionation factor calculated by Farkas and Farkas [16], namely 3·8. This factor is given by the equilibrium constant for the reaction

$$H_2O + HD = DOH + H_2, \quad K = 3{\cdot}8,$$

and the factor will be constant over the whole range of concentrations if the equilibrium constant for the reaction,

$$H_2O + D_2O = 2\,HDO,$$

is equal to 4. It is known to be 3·28 [10], and hence some variation with composition could be expected, though no such variation has been observed. Appleby and Ogden [17], from experiments designed to remove the gas from the electrode as rapidly as possible, to prevent exchange reactions taking place, found fractionation factors as high as 100, but Walton and Wolfenden [18] have not been able to confirm this. The experiments, as usually made, involve long periods of electrolysis during which the concentration of the residual hydrogen and deuterium is changed appreciably so that analyses can be made. During such periods of time it is entirely probable that electrodes change their character and exact results cannot be expected.

Eucken and Bratzler [19], Holleck [20] and Hutchison [21] have attempted the separation of the lithium isotopes by electrolysis, using a mercury cathode, and found that they are fractionated slightly. Eucken and Bratzler secured 1·07 as an upper limit, while Holleck and Hutchison secured 1·079 and 1·062 respectively for the fractionation factor.

Selwood, Taylor, Hipple and Bleakney [22] found that the oxygen of a 1-c.c. residue from the electrolysis of 117 litres of water contained 0·222 per cent of ^{18}O instead of 0·202 per cent as in the original water. Johnston [23] found that the fractionation factor is 1·008 with the heavier isotope preferentially remaining in the liquid.

The experiments on lithium and oxygen show that no effective separation of the isotopes of any element other than hydrogen can be expected by this method.

Horiuti and Okamoto [24] have found that electrode materials for the hydrogen-deuterium separation can be divided into two groups, one giving fractionation factors near 3 and the other near 6. The first group includes silver, nickel, platinum, gold, copper and lead (in alkaline solutions) and the second group mercury, tin and lead (in acid solutions). Walton and Wolfenden [25] have confirmed this work and found that the temperature coefficient of α is large and negative for the first group and nearly zero for the second group. The former authors have postulated different mechanisms for the discharge processes for these two groups, each being a two-step process, the first step being the same for both. These may be written as follows:

$$H_3O^+ + Me + e^- \rightarrow H_2O + MeH, \qquad \ldots.(2)$$

$$2MeH \rightarrow 2Me + H_2 \text{ (Group I)}, \qquad \ldots.(3)$$

$$MeH + H_3O^+ + e^- \rightarrow Me + H_2 + H_2O \text{ (Group II)}. \qquad \ldots.(4)$$

Okamoto, Horiuti and Hirota [26] have calculated the fractionation factor for a nickel electrode using the mechanism proposed by the transition state method, and secured 7·1 for the calculated fractionation factor, in excellent agreement with general experimental values.

Two experiments seem to be of especial significance in connexion with the theory of the process. The fact that with a mercury cathode the lithium isotopes are separated by electrolysis to an extent larger than can be accounted for on an equilibrium basis, shows that at least in this process it is the direct discharge of the ion which leads to separation. It does not necessarily follow

that this is also the case for hydrogen and deuterium, but the experiments by Farkas [27] show that hydrogen and deuterium are fractionated when the electrode is of palladium, in which the hydrogen is absorbed and not liberated as a gas. This shows that the recombination process is not the only effective one in this case, but that the initial step postulated by Horiuti and Okamota may be rate-determining as suggested by Gurney [28]. One may conclude that this separation process deserves more extensive experimental study and also that no extensive separation of other isotopes than hydrogen can be expected by the electrolytic method.

2. *Distillation*

Lindemann [29] in 1919 pointed out that a considerable separation of isotopes should be secured by distillation, providing there was no zero-point energy. His experiments on the evaporation of lead were negative, and hence led to the conclusion that the zero-point energy existed and that no separation of the isotopes could be expected by distillation methods.

The Debye theory of the solid state enables us to make a prediction in regard to the relative vapour pressures of isotopic substances in the solid state provided the solid behaves essentially as a monatomic solid. The theory shows that no separation can be expected if the heat capacity of the solid substance differs only slightly from its high-temperature value of $6R$. Thus, in order to secure a separation because of effects predicted by this theory, it is necessary to have a substance which sublimes at a temperature where the solid has not approached the classical heat capacity value, and presumably the same condition may be expected in the case of liquids. This will be true only in the case of helium, hydrogen and neon. Keesom and Van Dijk [30] in 1931 pointed out that a separation of the neon isotopes should be possible by distillation methods. The theory predicted, and their experiments were in accord with, a difference of 4 per cent in the vapour pressure of ^{20}Ne and ^{22}Ne. In 1935 Keesom, Van Dijk and Haantjes [31] succeeded in making substantial separations of the isotopes of neon, and secured 4 litres of neon gas of atomic weight 20·043, 135 litres of gas of atomic weight 20·785, and 4 litres of gas of atomic weight 21·157. The apparatus used consisted of a glass fractionation column 2 metres long with 60 plates and cooled to the boiling point of neon. Refluxing took place in much the same way as with a high-temperature fractionation column of the same kind.

Urey, Brickwedde and Murphy [32] also used distillation to concentrate the hydrogen isotopes initially. In this case also the Debye theory is applicable, since hydrogen behaves as a monatomic solid. The theory used by these authors was the same as that used by Lindemann and by Keesom and his co-workers, but in this case the theory predicts a ratio of the vapour pressures of H_2 and HD at the triple point of ordinary hydrogen of 2·42 [33]. Experiments on the vapour pressures of H_2 and pure HD show that the separation factor for these two varieties is 2·75. This should give a very effective separation of hydrogen and deuterium, and they estimated that in the simple distillation carried out by them, in which approximately 4000 grams of hydrogen were evaporated

to a few cubic centimetres, the ratio of the isotopes was changed by a factor of 30.*

Keesom, Van Dijk and Haantjes [31], and also Keesom and Van Dijk [34], have used this method to concentrate the isotopes of hydrogen, and have secured as much as 30 per cent of deuterium by this method. Scott and Brickwedde [35] used this method to produce nearly pure HD, which has been prepared only by this method [33].

The Debye theory does not predict any difference in the vapour pressure of isotopic molecules, such as H_2O and D_2O, at temperatures so high that the heat capacity has approached the classical heat capacity value of $6R$, but many examples of such differences have been observed. Lewis and Cornish [37], and Washburn, Smith and Frandsen [38], showed that hydrogen oxide and deuterium oxide have different vapour pressures. The former authors also reported differences in the vapour pressures of the $H_2{}^{16}O$ and $H_2{}^{18}O$ waters. Henriques and Cornish [39] attempted the separation of the chlorine isotopes by the distillation of methylene chloride, but were not successful.

In recent years several workers have attempted the concentration of isotopes of other elements than hydrogen by the use of distillation methods. Most of this work has dealt with the problem of securing exceedingly efficient fractionation columns, since the differences in vapour pressures of these substances are very small. Thus H_2O and D_2O differ in boiling points by about 1·4° and $H_2{}^{16}O$ and $H_2{}^{18}O$ by about one-tenth of this amount. The ratio of the vapour pressures of some of these substances as determined by Urey and Wahl is given in table 1. They determined these ratios by a simple distillation method, making use of the Rayleigh formula applicable to this case.

Table 1

T °c.	$p(H_2{}^{16}O)/p(H_2{}^{18}O)$
11·25	1·013
23·0	$1{\cdot}008_8$
35·6	$1{\cdot}008_2$
46·35	$1{\cdot}007_7$
100	1·003 (Extrapolated)

Wahl, Huffman and Hipple [40] found that there is a slight difference in vapour pressures of $^{14}NH_3$ and $^{15}NH_3$. Recently Thode [41], using ammonia containing 70 per cent ^{15}N, determined the ratio of the vapour pressures of the ammonias by direct comparison. Table 2 gives a summary of his results.

Slight differences in the vapour pressures of $CH_3{}^{16}OH$ and $CH_3{}^{18}OH$, the former having the higher vapour pressure, have been observed by Urey and his co-workers, but almost no difference in vapour pressures was found in the case of $^{12}CH_3OH$ and $^{13}CH_3OH$.

* They found an increase by a factor of approximately 5 in their experiments, but also they used hydrogen which was prepared electrolytically from new electrolytic cells, and hence their hydrogen gas contained considerably less heavy hydrogen than normal.

Table 2. *Vapour pressures of $^{14}NH_3$ and $^{15}NH_3$*

T °K.	$p^{14}NH_3$ (cm. of Hg)	$p(^{14}NH_3)-p(^{15}NH_3)$ (cm. of Hg)
197·75	5·488	0·0286
204·4	9·060	0·0423
207·1	10·904	0·0465
208·8	12·593	0·0521
217·2	21·570	0·0819
228·2	41·030	0·1307
228·4	41·550	0·1321
234·9	59·560	0·1606
239·4	75·340	0·1872

Klar and Krauss [42] have found that the oxygens $^{16}O_2$ and $^{16}O^{18}O$ have vapour pressures differing by 5 per cent, but no use has been made of this difference in separating isotopes.

Stedman [43] has developed a very effective fractionation column which consists of crumpled sheets of gauze filled into tubes in a regular fashion, and reports that the height of an equivalent theoretical plate for the distillation of normal and cyclo hexanes is 1·25 to 2·5 cm. By the use of carbon tetrachloride he secured a slight separation of chlorine isotopes and produced two samples having atomic weights differing from normal by $-0{\cdot}042$ and $+0{\cdot}006$ units. He estimates that the boiling-point differences were about 0·05° C. A smaller separation occurred when methyl chloride was distilled.

Pegram, Huffman and Urey [44] have developed a rotating fractionation column consisting of a 15 cm. steel tube with stationary cones attached to the sides and a rotating rod supporting rotating funnels placed alternately with the stationary ones. The liquid flows down a stationary cone and drops off the lower edge on to a rotating cone. Centrifugal force drives it up the cone and throws it off the edge of the rotating cone where it drains to the next stationary cone. In this way they were able to secure a very low height for an equivalent theoretical plate for the distillation of water,* the height being from 2·5 to 4 cm. per theoretical plate. Using a column 10·5 metres long containing 1240 cones they increased the concentration of the oxygen isotopes by a factor of 5, and produced 0·955 per cent ^{18}O water.

In all cases where exact calculations can be made, such as the vapour pressure of solids using the Debye theory or the equilibrium constants of exchange reactions, we find that all differences in properties of isotopic substances disappear when kT becomes large as compared with the separation of the quantized energy levels. An exact theory of the vapour-pressure differences discussed above cannot be given, but we may expect that such differences have their origin in those degrees of freedom for which the energy-level separa-

* Fenske has found that the height required for a theoretical plate was much greater when water rather than organic liquids was distilled.

tions are large compared to kT. These may be the torsional vibrational frequencies of the molecules in the crystal or the internal vibrational degrees of freedom of the molecules. Herzfeld and Teller [45] have discussed the problem of these differences in vapour pressures. The torsional vibrations should always lead to a higher vapour pressure for the lighter molecule, as is usually observed. If the vibrations are harmonic the ratio of vapour pressures is given by the formula

$$\frac{p'}{p''} = \Pi_i \left\{ 1 + \frac{1}{24} \left(\frac{h}{kT} \right)^2 (\nu_i'^2 - \nu_i''^2) \right\},$$

where the product is taken over the three torsional vibration frequencies, ν_i. The frequencies of the internal vibrations are usually less in condensed phases, and because of the changes of zero-point energy in going from the solid or liquid to the gas, this leads to a higher vapour pressure for the heavier constituents, since the energy necessary for this change is less in the case of the heavier molecules [46]. Finally, van der Waals forces may be slightly different because of differences in polarizabilities arising from different amplitudes of zero-point vibrations.

The theory of the distillation process has been discussed for many years, and it would seem that little new could be presented. The Rayleigh distillation formula for a simple-process distillation is given by equation (1), but simple distillation could not give an appreciable change in isotope concentrations; hence rectification or fractionation columns must be used.

Distillation has been used as a means of separating substances having considerable difference of boiling points, and it is only in recent years that fractionation columns have been developed which are capable of separating isotopes and other substances having nearly the same vapour pressures. The theory of such columns adapted to the separation of isotopes is simple in some respects in that the temperature may be assumed to be uniform throughout the column and that the heats of vaporization of the two constituents are nearly the same. The operation of these columns with the object of separating isotopes is also carried on under very different conditions from those obtaining in the separation of most chemical substances. In spite of all efforts to keep the hold-up (that is, the working material per unit length) down to a minimum, and to secure a high total flow through the column, long periods of time are required to reach a steady state. Also the materials used are not usually valuable and need not be conserved. Experiment shows that a fractionation column of the packed type gives concentrations in accord with the formula

$$\frac{N_t}{1-N_t} \Big/ \frac{N_b}{1-N_b} = \alpha^K, \qquad \ldots(5)$$

where N_t and N_b are the mole fractions of one constituent at the top and bottom, respectively, α is the ratio of vapour pressures * and K is by definition the number of theoretical plates, and experimentally is found to be proportional

* In the case of hydrogen and deuterium oxides α is the square root of the ratio of vapour pressures of H_2O and D_2O. See Lewis and Cornish [37].

to the length of the column. If α is large, as it is in many processes, and in particular in the cases of H_2 and HD, K may be small and yet a very large change in concentration may be secured. However, if α is nearly unity, as in the case of the separation of isotopes generally, so that the vapour pressures are nearly equal, K must be large in order to secure large changes in concentration. To secure large K, the length may be made large; increasing the surface of contact between gas and liquid by various types of packing is also effective in this respect. If the total hold-up of a given apparatus is known, and the ratio of concentrations at the top and bottom is known, it is a simple matter to calculate the total quantity of an isotope present in the column. Also, the transport of a constituent will be equal to the flow of liquid multiplied by the mole fraction of that constituent in the liquid minus the flow of gas multiplied by the mole fraction in the gas. When a steady state is reached this transport will be the same at every point in the column, and if no material is removed at either end will be zero.

If we assume that the velocity of exchange of two constituents between vapour and liquid takes place in accordance with an equation of the type

$$H_2{}^{16}O \text{ (liquid)} + H_2{}^{18}O \text{ (gas)} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} H_2{}^{18}O \text{ (liquid)} + H_2{}^{16}O \text{ (gas)},$$

and if we assume that diffusion along the length of the column is negligible, the differential equations for the distillation process are

$$H\frac{\partial N}{\partial t} - L\frac{\partial N}{\partial z} - k_1 N(1-n) + k_{-1} n(1-N) = 0, \qquad \ldots.(6\,a)$$

$$h\frac{\partial n}{\partial t} + l\frac{\partial n}{\partial z} + k_1 N(1-n) - k_{-1} n(1-N) = 0, \qquad \ldots.(6\,b)$$

where $H(h)$ is the hold-up of liquid (gas) per unit length, $L(l)$ the flow per unit time of liquid (gas), k_1 and k_{-1} are the velocity constants for the two reactions indicated in the reaction above and n and N the mole fractions of $H_2{}^{18}O$ in the gas and liquid phases, respectively.

The general integral of these equations has not been found* though the integral for the steady state, $\partial N/\partial t = \partial n/\partial t = 0$, can be secured quite easily. In the special case where there is no forward flow removed from the column, so that L and l are equal, N must be equal to n and either equation integrates to give

$$\ln\left(\frac{N_t}{1-N_t}\Big/\frac{N_b}{1-N_b}\right) = (k_1 - k_{-1})Z/L, \qquad \ldots.(7)$$

where N_t and N_b are the mole fractions of one constituent ($H_2{}^{18}O$ in our illustration) at the top and bottom respectively, and Z is the length of the column. This is the observed equation (5), if

$$K = \frac{(k_1 - k_{-1})Z}{L \ln \alpha}. \qquad \ldots.(8)$$

*Dr. George Kimball of Columbia University has found a solution of the equations, but it proves to be of quite impractical complexity.

In this case the column would be operated without loss of material either at the top or bottom, i.e. with total reflux.

When the column is operated with forward flow, i.e. $L \neq l$, the equations can be integrated for the case of the steady state. If we assume a large body of liquid at the bottom of the column whose composition changes very slowly, and a forward flow at the top equal to $(l-L)$, the transport of one constituent at any point in the column must be equal to the loss of this constituent at the top,

$$ln - LN = (l-L)N_t. \qquad \ldots.(9)$$

With this equation n can be eliminated from either of the equations (6), the derivative with respect to the time set equal to zero and the equation integrated. The integral is somewhat involved but can be used to show the relationship between the rate of production and the overall fractionation factor for any particular operation. At the beginning of the operation of the column under total reflux the transport is

$$T = L(n-N) = LN(\alpha-1)\frac{(1-N)}{1+(\alpha-1)N}, \qquad \ldots.(10)$$

so that the transport is proportional to N when N is small and to $(1-N)$ when N is nearly unity, and has a maximum value when $N = 1/2$ and α is nearly unity. This relation is qualitatively the same also for the exchange reaction and thermal diffusion methods to be discussed. In comparing the rates of production by various methods, one must remember that the rate of production is about proportional to the mole fraction, N, of the constituent in the starting material if this is small, or to $(1-N)$ in the concentrated material if this is small.

To secure a working formula for the time-dependence of the concentration, Huffman and Urey assumed for their particular case (total reflux with a large reservoir at one end and a very small one at the other, the total reflux end), that the following conditions held: a fraction of the column near the total reflux end, either top or bottom, has a distribution of its constituents in accordance with formula (5), and the remainder of the column has no gradient of concentration at all. It is then possible to calculate the total amount of material in the column at any time. Also, the rate at which this increases with the time is equal to $L(n-N)$ or to $LN(\alpha-1)$, if N is small and α nearly equal to unity (equation (10)), since this is the transport in the part of the column where there is no concentration gradient. In this way they secured the equation

$$N_b \ln\frac{N_b}{N_t} + (1-N_b)\ln\frac{1-N_b}{1-N_t} = \frac{LN_b(\alpha-1)K\ln\alpha}{H}t \qquad \ldots.(11)$$

for the relation between the concentration N_t at the reflux end of the column and the time. The conditions for rapid increase of concentration are easily seen from this formula:—a large number of plates, K; a large value of the fractionation factor, α; a small value for the hold-up, H; rapid total flow, L; and as high an initial concentration N_b as possible. If total reflux occurs at

the bottom and there is a large reservoir at the top, we interchange N_b and N_t. In the former case N_b decreases with the time and in the latter case it increases. They found that this formula agreed with their experiments on the concentration of the oxygen isotopes.

In Huffman and Urey's most satisfactory experiment, water was pumped into the top of the fractionation column at a uniform rate of 33 c.c. per minute. At the bottom it was boiled by means of a flash boiler heated with steam. The vapour leaving the top was condensed and run to waste. In this case total reflux occurred at the bottom. The column was 10·5 metres long and contained 1240 cones of the type described above, and had a total hold-up of 4300 c.c. of water. The simple process fractionation factor was about 1·006. A change of five times the normal concentration of ^{18}O was secured, requiring about 250 theoretical plates. With the use of the formulae and the rate of increase with time, the column was estimated to have 425 theoretical plates, but would have required 50 days of steady operation to produce 2·5 per cent $H_2{}^{18}O$ water, at the rate of 0·67 of a gram of pure $H_2{}^{18}O$ water per 24 hours.

The formulae developed in this section are also applicable to the exchange reaction method of separating isotopes, and formula (11) has been found to hold in these cases as well.

High concentrations of ^{18}O have not been secured by this method, largely because no one has taken the time to do so. Undoubtedly a cascade arrangement of columns similar to those used by Urey and his co-workers for exchange reactions and described in the next section should be used. Though the Pegram column is very efficient, it seems to the writer that Stedman or Fenske columns would probably be better for laboratory experiments. The column recently described by Goldwasser and Taylor [47] would probably be best for the smallest unit in the cascade.

Distillation does not seem to be a promising method for the separation of other isotopes. Nitrogen isotopes might be separated in this way, but maintaining long columns at or below the boiling point of ammonia would be difficult. Similar difficulties exist in the case of the distillation of oxygen. It might be possible to work on boron using one of its volatile compounds, but as yet no definite knowledge exists in regard to differences in vapour pressure.

3. *Exchange reaction methods*

That the equilibrium constants for exchange reaction methods were not unity in all cases was shown first by Urey and Rittenberg [48] from theoretical calculations which were later verified experimentally. The success in separating the hydrogen isotopes led to the study of such reactions, and led naturally to the calculation of equilibrium constants for exchange reactions involving other elements than hydrogen. These calculations as carried out by Urey and Greiff [49] showed that the equilibrium constants for such exchange reactions differed slightly from unity, and showed further that when this is the case, the isotopic composition of two chemical compounds which are in equilibrium with respect to their isotopes will not be the same, and hence that a slight fractionation of the isotopes will occur in such reactions.

Contrary to the case of separation of isotopes by electrolysis and by distillation, the theory of this method was well understood at the time that attempts to separate isotopes by such methods were made. Table 3 gives the results of such theoretical calculations for a number of chemical reactions as given by Urey and Greiff. In particular it should be noted that lithium, carbon, nitrogen, oxygen exchange reactions have equilibrium constants differing from unity by a few per cent, and that as the mass of the atom increases, for instance in the cases of chlorine and bromine, the deviation of these constants from unity rapidly decreases. Moreover, as the temperature increases the deviation from unity decreases rapidly but not monotonically in all cases. The bromine reaction constant changes from greater than unity to less than unity, a general effect which was first pointed out by Urey and Rittenberg [48] for the hydrogen-hydrogen iodide reaction, and which has been more carefully studied by Murphy [50]. If these differences are to be utilized for the concentration of isotopes, it is necessary in general to use low temperatures and to work on the elements of low atomic weight.

Table 3

Reaction	Equilibrium constants		
	273·1° K.	298·1° K.	600° K.
$\frac{1}{2}S^{16}O_2+H_2{}^{18}O=\frac{1}{2}S^{18}O_2+H_2{}^{16}O$	1·034	1·028	1·003
$\frac{1}{2}C^{16}O_2+H_2{}^{18}O=\frac{1}{2}C^{18}O_2+H_2{}^{16}O$	1·064	1·054	1·014
$\frac{1}{2}{}^{16}O_2+H_2{}^{18}O=\frac{1}{2}{}^{18}O_2+H_2{}^{16}O$	1·024	1·020	1·003
$^{13}CO+{}^{12}CO_2={}^{12}CO+{}^{13}CO_2$	1·098	1·086	1·029
$\frac{1}{2}{}^{36}Cl_2+H^{37}Cl=\frac{1}{2}{}^{37}Cl_2+H^{35}Cl$	1·004	1·003	1·00015
$\frac{1}{2}{}^{79}Br_2+H^{81}Br=\frac{1}{2}{}^{81}Br_2+H^{79}Br$	1·0005	1·0004	0·99997
$\frac{1}{2}{}^{14}N_2+{}^{15}NO=\frac{1}{2}{}^{15}N_2+{}^{14}NO$	1·016	1·015	1·007
$^6LiH+{}^7Li={}^7LiH+{}^6Li$	1·028	1·025	1·008

In order to multiply the smaller differences in chemical properties it is necessary to make use of two-phase counter-current methods, the most convenient of which is very similar to the distillation process described above. Urey and his co-workers have used gas-liquid equilibria and solid-liquid equilibria, while Lewis and MacDonald have used liquid-liquid equilibria.

The methods which have been successfully used for concentrating the isotopes of nitrogen and carbon make use of the following exchange reactions :

$$^{15}NH_3(g)+{}^{14}NH_4{}^+(aq)={}^{14}NH_3(g)+{}^{15}NH_4{}^+(aq) \quad \ldots.(12)$$

and

$$H^{12}CN(g)+{}^{13}CN^-(aq)=H^{13}CN(g)+{}^{12}CN^-(aq). \quad \ldots.(13)$$

The counter-current system for the first of these reactions consists of an ammonium salt in water solution, flowing down through a fractionation column, the addition of sodium hydroxide to this solution at the bottom, and removal of all the dissolved ammonia by boiling and passage of the

ammonia upward through the fractionation column. The equilibrium constant for this reaction can be written

$$\frac{[^{15}NH_4^+]}{[^{14}NH_4^+]} \Big/ \frac{[^{15}NH_3]}{[^{14}NH_3]} = K. \qquad \ldots.(14)$$

As a matter of fact, the equilibrium constant for the ammonium-ion exchange reaction is approximately 1·023, and hence a higher concentration of ^{15}N will be found in the solution. The concentrations expected in this case can be calculated by the use of equation (5), and the transfer which can be expected could be calculated from equation (10). Concentration of ^{15}N will occur at the bottom of the column and in this case total reflux at the bottom should be used.

In the case of the hydrogen cyanide-cyanide exchange reaction, the equilibrium constant is approximately 1·015, and this results in an increased concentration of ^{13}C in the gas phase. In this case, hydrogen cyanide gas is fed into the bottom of the column. At the top it is absorbed in sodium hydroxide, and the sodium cyanide solution so formed flows down the column. Concentration of ^{13}C will occur at the top of the column, and reflux at the top should be used in this case.

Urey and Aten [51] first investigated the possibility of using exchange reactions as a means of separating the nitrogen isotopes. They found that little fractionation was secured between ammonia gas and ammonium hydroxide solution. They then investigated the exchange reaction between ammonium gas and ammonium ion in the form of its chloride by adding sodium hydroxide to a solution of ammonium chloride in small amounts, and removing the ammonia by a current of air after each addition. In this way they changed the ratio of the isotopes by about 10 per cent, and by using the Rayleigh formula, equation (1), secured 1·027 for the simple process fractionation factor, that is the equilibrium constant of the exchange reaction.

Urey, Huffman, Thode and Fox [52] used the fractionation column used previously for the distillation of water, to investigate again some of the nitrogen exchange reactions of ammonia. In addition to the ammonium hydroxide reaction they also tried to separate the isotopes by counter-current scrubbing of ammonia against a solution of ammonia in alcohol, and found only slight separation. They found that the best separation was secured by the use of ammonia gas and ammonium nitrate solution, but, due to the corrosive properties of the nitrate, they used ammonium sulphate to produce an appreciable concentration of ^{15}N by these counter-current methods. The final concentration secured was 2·5 per cent instead of the natural abundance of 0·38 per cent.

Using equation (11) it is possible to show that no very high concentration of ^{15}N could be secured by the use of their fractionation apparatus because of the very long time which would be required for the steady state to be reached. Thode, Gorham and Urey [53] designed a cascade system of three units to produce a high concentration of ^{15}N by these methods in a reasonable length of time. Figure 1 shows the flow system used by them. Ammonium nitrate solution was passed into the first unit at F, which changed the

ratio of the isotope by a factor of approximately 7. Approximately sixth-sevenths of the flow was diverted to an apparatus S, where sodium hydroxide was added, and the solution boiled to remove the ammonia which was sent back through E. One-sixth of the flow at the bottom of the unit E passed to the top of the unit E' which changed the ratio of the isotopes by a factor of 9. Approximately eight-ninths of the solution passed to the apparatus S, where sodium hydroxide was added, the ammonia removed and passed upward through E', while one-ninth of the solution passed to unit E''. Unit E'' changed the ratio of the isotopes by a factor of 11; all of the ammonia was removed in the apparatus S'' and passed back through the whole system.

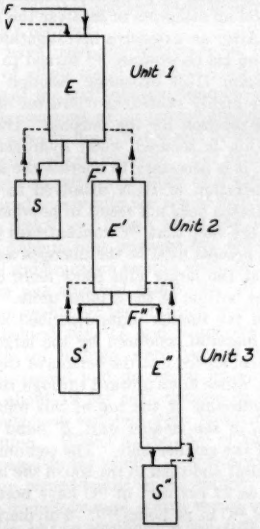


Figure 1. Three-unit cascade system.

Since the transport of each of these units will be approximately proportional to the concentration of ^{15}N (equation (10)), the transport will be approximately the same in each of these units, but since the units are smaller the hold-up will be smaller, and hence the time required to attain equilibrium will be shorter. The actual apparatus consisted of glass tubes, the large unit consisting of 24 metres of glass column, 7cm. inside diameter, packed with $\frac{1}{4}$-inch Berl saddles. The second consisted of 12 metres of glass column, 2·2 cm. inside diameter, packed with 6 mm. one-turn helices of the Fenske [54] type. The third unit consisted of 7·5 metres of glass column, 1 cm. inside diameter, packed with

5 mm. one-turn glass helices. The units were all mounted in sections with pumps to transfer the liquid from the bottom of one section to the next. S, S' and S'' are refluxing fractionation columns to remove the ammonia from the water solutions. Using this apparatus the maximum concentration of 72·8 per cent ^{15}N was secured, and approximately 31 gm. of ^{15}N was transferred from ordinary nitrogen to samples ranging in concentration from 0·57 to 70·6 per cent. The transport of their apparatus was about 0·75 grams per day.

The most attractive reaction to use for concentrating carbon would be the exchange reaction between CO_2 and bicarbonate ion. Unfortunately this reaction is very slow and all attempts to catalyze the reaction in a satisfactory way have failed [55]. After an extensive investigation of other reactions that might be used, Urey and his co-workers [56] turned to the use of the hydrogen cyanide-cyanide (equation (13)) exchange reaction for carbon mentioned above. Because of the highly poisonous character of hydrogen cyanide, this is not a very attractive reaction for the purpose. Owing to the fact that no appreciable concentration is secured when hydrogen cyanide is exchanged with its water solution, it is necessary to operate the apparatus under vacuum to decrease the concentration of HCN dissolved in the water solution and thus a leak in the apparatus does not result in poisonous gas escaping into the room. Owing also to the fact that ^{13}C concentrates in the gas, it is necessary to operate the cascade process used in the nitrogen separation with the small units at the top so that the larger unit feeds more concentrated samples of hydrogen cyanide to the bottom of the smaller units. The apparatus used for this purpose consists of the smaller units described above, with the smallest unit operating on the material produced by the larger of the two. Sodium cyanide and sulphuric acid are fed into the bottom of the larger unit. Hydrogen cyanide gas is liberated, which flows upward through the column E', is partially absorbed by sodium hydroxide at the top of this unit, while part of the gas flows on to the bottom of the smaller unit, E'', and is absorbed by sodium hydroxide at the top of the smaller unit. The sodium cyanide formed in this flows down the smaller unit and then to the top of the larger unit. In this way concentrations as high as 22 per cent of ^{13}C have been produced at a rate of approximately 0·1 gr. of ^{13}C in 24 hours [57]. Polymerization of the hydrogen cyanide in the alkaline sodium cyanide solution is greatly reduced by the addition of sodium sulphide to the solutions and does not cause serious difficulty. No difficulty because of the poisonous hydrogen cyanide has been encountered even when the glass tube of a refluxing column broke off completely. Canaries were kept in the room as indicators of hydrogen cyanide, but no canary was killed in these runs.

Thode, Gorham and Urey [58] have reported that ^{34}S is concentrated by counter-current scrubbing of sulphur dioxide against a bisulphite solution, the heavier isotope concentrating in the sulphite solution. The concentration of ^{34}S was changed by a factor of about 3, when the smallest unit described above was used.

Lewis and MacDonald [58] used the exchange reaction between lithium

dissolved in mercury and lithium ion dissolved in an alcohol solution for the fractionation of the lithium isotopes. A vertical tube 18 metres long was filled with a solution of lithium chloride in anhydrous methyl alcohol, and a lithium amalgam flowed in small droplets down through this solution. At the bottom, the lithium of the amalgam was converted to lithium chloride dissolved in alcohol, and then fed into the bottom of the column thus maintaining a counter-current flow with total reflux at the bottom. In this way they were able to produce lithium containing 16·3 per cent of the heavier isotope, instead of the 8 per cent in natural lithium. The heavier isotope in this case concentrates in the amalgam. In the second run they used lithium bromide dissolved in one part of ethyl alcohol and two parts of dioxane. The percentage of ^{7}Li was increased to 13·2 per cent.

Taylor and Urey [59] have used the difference in the exchange reaction between water solutions of alkali metal ions and solid zeolites as a means of concentrating the isotopes of lithium, nitrogen (ammonium ion) and potassium. A true solid-liquid counter-current arrangement analogous to the gas-liquid and liquid-liquid systems already described would be difficult to devise. They used the methods of chromatophoric analysis. For example, a tube was filled with a sodium zeolite covered with a solution of sodium chloride. Lithium chloride solution was poured slowly into the top of this tube and displaced the sodium chloride solution which was drawn off at the bottom as the solution entered at the top. The lithium ions replaced the sodium ion in the zeolite, but the ^{6}Li displaced the sodium more readily than did the ^{7}Li and hence the leading sample of lithium was enriched in the ^{7}Li. If then the lithium was displaced from the lithium zeolite by the addition of sodium chloride solution, the ^{6}Li isotope trailed behind the ^{7}Li and hence was concentrated in the trailing sample. Samples of lithium were secured in which the heavier isotope of lithium was increased and decreased by about 25 per cent of its natural abundance, and about 10 per cent changes in the normal abundance of the potassium and nitrogen isotopes were secured, the latter by using ammonium chloride solution. Whilst in the case of lithium the light isotope is preferentially concentrated in the zeolite, in potassium and nitrogen it is the heavier isotopes which are concentrated in the zeolite. Hence it would seem that the process is not a kinetic process, which would always concentrate the light isotope in the zeolite, but an equilibrium one, just as the other exchange reactions discussed above. This method might be developed into an important one for producing increased concentrations of ^{7}Li, but hardly for producing nearly completely separated isotopes.

4. *Separation by diffusion*

In 1932 Hertz [60] described a cascade diffusion apparatus (figure 2) with which he was highly successful in concentrating the isotope of neon of mass 22. The operation of this apparatus can be illustrated by considering the gas entering the diffusion unit *R* through the tube *A*. A fraction of the gas diffuses through the tube *R*, and leaves through *B* to the following diffusing unit, as shown by the arrows. The remaining fraction of the gas, *l–f*, enters the diffusion

unit S and part of the gas diffuses into C and is pumped from the diffusing unit S to the diffusing unit R by the pump P_3. The final gas leaves through D to the preceding diffusion unit. In this way the light gas flows to the right and the heavy gas to the left. Each unit produces a certain fractionation factor, say s, i.e. the ratio of isotopes at one unit divided by this ratio at the preceding unit, while a system of n such units will give an overall fractionation factor of s^n. At the extreme left is a volume containing the heavier fraction and at the right a volume containing the lighter fraction. Mercury diffusion

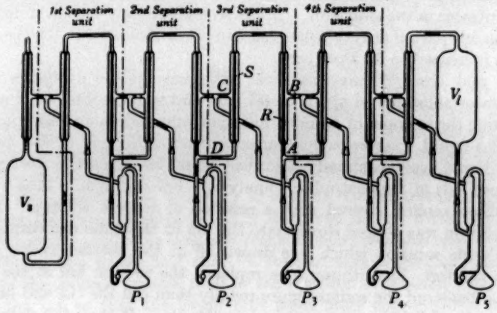


Figure 2. Hertz's diffusion apparatus.

pumps are used to cause a circulation of the gas, as shown. The volumes for the heavy gas and the light gas can be adjusted, and in general if one wishes to produce a high concentration of the light gas it is necessary to make the V_s large and the V_l small, and the reverse if one wishes to produce a high concentration of the heavy gas. Hertz's original apparatus consisted of 24 pumps and 48 diffusing tubes. By means of this apparatus it was possible to change the ratio of the neon isotopes by a factor of 8 in a single operation, and by repeating the process a large separation could be secured.

Hertz [61] derived an approximate theory showing that the fractionation factor for one unit of this apparatus should be given by

$$q = \frac{f^\mu}{1-(1-f)^\mu}, \qquad \ldots.(15)$$

where f is the fraction of the gas entering at A, which diffuses through R. In the experiments which he described he made this equal to 1/3, and also the fraction diffusing through the tube S equal to 1/3. μ is the square root of the ratio of the masses.

Harmsen [62] has described a more effective apparatus of this kind, which changed the ratio of the neon isotopes in a single operation by a factor of approximately 40, and by repeated operation produced a sample of ^{22}Ne containing only about 1 per cent of ^{20}Ne. The volumes at the light and heavy ends of the apparatus in one experiment were about 30 and 5 litres respectively.

The pressures at the two ends were 11 and 8·5 mm. of mercury. After 12 hours the ratio of the isotopes was 2·5 to 1 instead of 9 to 1. This produced an approximately 30 per cent sample of ^{22}Ne and added approximately 10 c.c. (S. T. P.) of ^{22}Ne to the more concentrated material in 12 hours.

Harmsen, Hertz and Schutze [63] doubled the number of pumps and the number of diffusing units so that in one operation nearly pure ^{22}Ne could be produced. In approximately 8 hours this apparatus was able to produce spectroscopically pure deuterium from a 10 per cent sample of hydrogen-deuterium mixture. In order to do so it was necessary to introduce a discharge tube toward the middle of the apparatus to establish equilibrium between HD, H_2 and D_2, thus permitting final concentration of the pure deuterium molecule. Wooldridge and Smythe [61] and Wooldridge and Jenkins [65] used this method to produce 16 per cent ^{13}C and 3 per cent ^{15}N and made some study of the efficiency of various types of apparatus.

In 1922 Hertz [66] proposed a diffusion apparatus in which diffusion took place through streaming mercury vapour, but no serious development of this

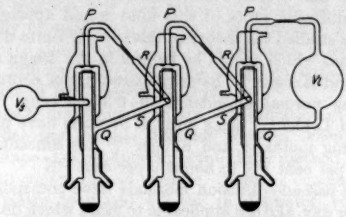


Figure 3. Hertz's vapour diffusion apparatus.

method took place during the next ten years. MacGillavry [67] then constructed an apparatus quite similar to that used by Hertz, and used it to produce some separation of the hydrogen isotopes, securing a fractionation factor as high as 2·3 for a single diffusion unit. MacGillavry also discussed the theory of this diffusion process, deriving theoretical expressions for the fractionation factor and the rate of transport through the unit.

Hertz [68] has modified his original apparatus for diffusion through streaming vapours and arranged these in a cascade similar to the cascade used for the diffusion of isotopes through porous walls. Figure 3 shows cascades of three units of this kind. The mercury is boiled through the inner tube and condenses in the top part of the apparatus where there is a water-jacketed cooling surface. Part of the mercury vapour passes into tube *P* and part condenses in the jacketed chamber. The gas diffuses from this chamber through the streaming mercury vapour and passes partly into the tube *P*. That which passes through this tube proceeds through the capillary *R* into the next diffusing unit. This is a light fraction. The heavy material that remains behind flows through *S* to the preceding unit, while the light fraction from this unit enters

through R. In this way the diffusing material of the second unit is constantly replenished from the light fraction from the first unit, while the light fraction of the second unit is passed to the third diffusing unit, and a counter-current flow is maintained in this way. A heavy reservoir is placed at the left end of the apparatus, indicated by V_h, and a light reservoir at the right side indicated by V_l. Using a set of twelve such units, Hertz succeeded in changing the ratio of the Ne isotopes by a factor of 10 in 45 minutes.

Barwich [69] has studied the separation by units of this type, studying such mixtures as helium-hydrogen, ^{20}Ne–^{22}Ne, argon-air, neon-helium and argon-helium, and has derived a theoretical equation for the fractionation factor of a single unit. Barwich found fairly good corroboration of this theory, but Sherr [70] was unable to find very good agreement between the theory and the experiment. Sherr studied the separation of the bromine isotopes using methyl bromide, the carbon isotopes using CH_4, the oxygen isotopes using oxygen gas and water vapour, the neon isotopes, the argon isotopes, and the separation of argon-carbon dioxide mixtures. He further gave a theory of the effect of many units upon the overall fractionation factor, as well as the rate at which a series of diffusing units of this kind would approach a steady state. Blumenthal [71] has studied the theory of this type of Hertz diffuser in detail and has secured solutions which should be helpful in the design of such apparatus.

Kopfermann and Kruger [72] have prepared argon containing 50 per cent ^{36}A instead of the normal abundance of 0·3 per cent. Capron, Delfosse, de Hemptinne and Taylor [73] using a set of 51 units have produced 30–32 per cent ^{13}C by diffusing methane, and Kruger [74] has concentrated a sample of nitrogen from 1·9 per cent ^{15}N to about 9 per cent ^{15}N.

These diffusion-cascade-diffusion methods have transports of about 1 c.c. (S. T. P.) of gas per day, and are applicable to gases which do not react with the mercury of the pumps and are not decomposed by mercury boiling under reduced pressure.

5. *Thermal diffusion*

Clusius and Dickel [75] in 1938 introduced a method for the separation of isotopes by the use of thermal diffusion, which has proved to be exceedingly effective in many cases. The principle of the method can be described as follows: If there is placed between two vertical parallel walls, one hot and the other cold, a mixture of two gases of different molecular weights, separation of the two constituents will occur because of the following effects: (1) due to thermal diffusion the resulting concentration of the heavy constituent will be higher on the cold wall and that of the lighter constituent will be greater on the hot wall; (2) thermal convection will occur, there being a flow of the cold gas downward near the cold wall and of hot gas upward at the hot wall.

These two effects will result in two counter-current streams of different concentration in the two constituents, resulting in an increased concentration of the heavy constituent at the bottom and of the light constituent at the top of the apparatus. This process will continue until back diffusion along the tube mixes the gases as rapidly as a combination of thermal diffusion and convection

separates them. The apparatus used by Clusius and Dickel consisted of a vertical tube with a heated wire stretched through the centre. In their first experiments they used tubes varying from 65 cm. to 2·9 metres in length, and a temperature difference between the wire and the walls of about 600°. Mixtures of gases such as bromine and helium were completely separated, pure helium appearing at the top and pure bromine (liquid) at the bottom. Carbon dioxide was completely separated from hydrogen-carbon dioxide mixtures. The proportions of the constituents of air were changed in a marked degree. Neon of atomic weight 20·62 instead of 20·18 was produced and the proportions of ^{37}Cl and ^{35}Cl were changed from the normal ratio of 23 to 77 to a ratio of 40 to 60. Later these authors [76] reported that they had succeeded in producing 99·6 per cent ^{35}Cl and 99·4 per cent ^{37}Cl. They used a 20-metre tube for the production of ^{35}Cl, and produced 28 c.c. and 16 c.c. samples of HCl gas, whose chlorine had atomic weights 35·021 and 34·979 respectively. Using a 36-metre tube they were able to produce 8 c.c. of HCl gas containing 99·4 per cent ^{37}Cl per 24 hours. This chlorine had atomic weight 36·956. These experiments resulted in a practically complete separation of the isotopes of chlorine, and thus a problem, whose solution was first attempted some 25 years ago by other workers, was finally brought to a successful conclusion. The transport given above would indicate about 6 mg. of ^{13}C per 24 hours, if equation (10) is used to adjust for the different concentrations of ^{13}C in natural carbon and the 99·4 per cent sample of $H^{37}Cl$ reported above.

Groth [77] has made experiments on the concentration of the xenon isotopes and has shown that the thermal diffusion method when applied to xenon has about the same speed of production as the Hertz method, but that it is much more convenient to construct and operate. Thus he concludes that a thermal diffusion tube 1 metre long and 0·5 cm. in diameter, with a central wire heated to between 1200 and 1650° C. is about equivalent to 12 Hertz diffusion pumps so far as the concentration is concerned, and with a rate of production about the same in both cases. He found that when a tube 1·2 cm. in diameter was used, the maximum separation increased until the wire was at about 1000° C., after which the concentration decreased as the temperature of the wire increased. When a tube 0·5 cm. in diameter was used, no such maximum effect was observed. Using a tube 2·5 metres in length and 0·5 cm. internal diameter, he was able to change the atomic weight of two xenon fractions by 1·57 atomic weight units. (Groth in the text gives 1·57 atomic weight units, but the figure which he publishes indicates one unit only.)

Bramley and Brewer [78] have used an apparatus consisting of two concentric glass tubes, the outer one cooled by water and the inner heated by electric current. In this way the separation chamber is increased in size though the temperature differences which can be maintained in a glass apparatus are much less. Working on ammonia-methane mixtures they secured a fractionation factor of 1·3 for a 1-metre tube and of 2·0 for a 3-metre tube. The separation factors secured in this way are not as high as those reported by Clusius and Dickel, probably because of the smaller temperature differences used.

Korshing and Wirtz [79], as well as Clusius and Dickel [80], have found that

an appreciable separation of the constituents of liquid solutions can be made by these methods. The first authors used a hot wire suspended in the middle of a glass tube as one apparatus. In a second apparatus the hot wire was replaced by a copper tube through which flowed hot oil. They used tubes of different lengths and applied them to mixtures of normal hexane and carbon tetrachloride and benzene and chloro-benzene. The degree of separation can be judged from one of their experiments in which a mixture of equal parts of *n*-hexane and carbon tetrachloride was changed in 9 hours to 70 per cent CCl_4 at the top and 26 per cent CCl_4 at the bottom. The tube was 60 cm. long and had a 0·35 mm. wall separation with a temperature difference of 45° C.

Clusius and Dickel used a tin apparatus, oblong in cross-section, 20 mm. wide, 1 mm. thick and 1·5 metres long. They arranged this so that the one side could be heated with steam and the other cooled with cold water, and arranged so that it could be inclined to give an optimum value for the separation. They found that when a 0·1 normal solution of sodium chloride was placed in this apparatus, with a temperature difference of about 80°, the concentration of the sodium chloride increased by a factor of 3·6 at the bottom. In the case of a mixture of acetone and water containing 42·2 mole per cent of acetone, the concentration of acetone changed to 6·2 mole per cent at the bottom after six hours, with a temperature difference of 40°. They also observed a small change after eight hours, in the ratio of H to D in a mixture of hydrogen and deuterium oxides, employing a temperature difference of 80° C. Korshing and Wirtz [81] continued their investigations and found that the separation increases with the time at first very rapidly, and later at a rate that is about proportional to the square root of the time. They expect that it will require periods of the order of 1000 days for an approach to equilibrium since the coefficients of diffusion are very small. They made fairly extensive investigations of the sedimentation of salts by this method. Some of their results are given in the following table :—

Table 4

Exp. no.	Substance	Conc. in mol/litre	Time in hours	Length of tube	$\frac{n\,(\text{top})}{n\,(\text{bottom})}$
2	$ZnCl_2$	0·1	72	90 cm.	1·75
3	$ZnSO_4$	0·1	48	90	4·4
4	,,	0·4	48	90	7·3
5	,,	0·6	48	90	9·5
7	NaCl	0·5	48	90	1·8
8	Na_2SO_4	0·5	48	90	7·3

It will be noted that zinc sulphate and sodium sulphate concentrate much more rapidly than do zinc chloride and sodium chloride. They investigated the composition of the zinc isotopes secured in their experiment no. 4, and found proportions for the isotopes as given in table 5.

Table 5

	Normal zinc	Ratio at top	Ratio at bottom
$\frac{^{64}Zn}{^{68}Zn}$	2·93	3·2	2·7
$\frac{^{64}Zn}{^{66}Zn}$	1·89	1·95	1·7
$\frac{^{66}Zn}{^{68}Zn}$	1·56	1·65	1·56

The analyses were made spectroscopically and differ slightly from the accepted mass-spectrograph ratios. A very distinct separation is secured in a comparatively short time by this method. These authors have further investigated the concentration of heavy molecules from water solution using Sudan I, which was found to settle rapidly, and chlorophyll which gave indecisive results.

The theory of the thermo-diffusive separation of isotopes in gases has been discussed by Waldmann [82], van der Grinten [83], and particularly by Furry, Jones and Onsager [84]. The latter authors have made a very comprehensive study of this problem. The differential equation is

$$c(\mathbf{v}_1-\mathbf{v})=-D_{12}\operatorname{grad} c_1+(D_T/T)\operatorname{grad} T, \qquad \ldots.(16)$$

where c_1 is the relative particle density of one constituent with $(c_1+c_2)=1$, $\mathbf{v}_1$ is the convective velocity of the lighter constituent and $\mathbf{v}$ is the convective velocity of the gas as a whole. D_{12} is the coefficient of ordinary diffusion which in the case of isotopes can be set equal to the coefficient of self-diffusion, and D_T is the coefficient of thermal diffusion, which according to Enskog's theory is equal to $\alpha c_1 c_2$, where α is to a first approximation,

$$\alpha=\frac{105}{118}\,\frac{m_2-m_1}{m_2+m_1}. \qquad \ldots.(17)$$

The equations given by Waldmann are first approximations to the more exact results given by Furry, Jones and Onsager. The theory as developed by all these men is for two perpendicular parallel plane surfaces, and not for two concentric tubes, as is the case in most of the experiments. The equation for the transport of one constituent is

$$\tau_1=H\left\{c_1(1-c_1)-(2A_d)^{-1}\frac{\partial c_1}{\partial z}\right\}, \qquad \ldots.(18)$$

where

$$H=\frac{w^3\rho^2\varkappa g B}{6}\left(\frac{\Delta T}{T}\right)^2\sum_k\left(\frac{\Delta T}{2T}\right)^{2k}\frac{1}{(2k+1)(2k+3)(2k+5)}$$

$$=\frac{w^3\rho^2\varkappa g B}{6}\left(\frac{\Delta T}{T}\right)^2\left[1+\frac{1}{28}\left(\frac{\Delta T}{T}\right)^2+\ldots\right], \qquad \ldots.(19)$$

and

$$A_d=\frac{A}{1+K_d A},$$

$$A=\frac{63}{4}\frac{\alpha\eta D}{\rho g w^4}\left(1+\frac{1}{28}\left(\frac{\Delta T}{T}\right)^2+\ldots\right),$$

$$\frac{K_d}{K}=1890(T_1^2+T_1T_2+T_2^2)\left(\frac{D\eta}{w^3\rho g\Delta T}\right)^2,$$

where $2w$ is the distance between the surfaces and B their width, ρ the density, η the viscosity, both at the mean temperature T, and g the acceleration of gravity. In the case where the transport is equal to zero, equation (18) can be integrated to give

$$\frac{\kappa_1^e}{1-\kappa_1^e}\Big/\frac{c_1^0}{1-c_1^0}=e^{4Al} \qquad \ldots.(20)$$

with $2l$ equal to the length of the apparatus, and κ_1^e and c_1^0 the concentrations at the two ends of the column. The equation is similar to equation (7) for the case of distillation given above. If the transport is not equal to zero, but to a constant, the integral of equation (18) is

$$c_1^1=\frac{1}{2}-\frac{q}{2}\frac{q\tanh(2qAl-q_0)}{q_0\tanh(2qAl-q)}, \qquad \ldots.(21)$$

with $q=\left(1-\frac{4\tau_1}{H}\right)$ and $q_0=1-2c_1^0$, and c_1^1 and c_1^0 are the concentrations at the ends of the column. They also secure an approximate equation for the relaxation time, t_r, if the reservoir at the bottom is very large,

$$t_r=\frac{m_t}{H(1-2c_1^1)}(e^{4Al(1-2c_1^1)}-1), \qquad \ldots.(22)$$

where m_t is the mass of gas in the top reservoir and c_1^1 is about the mean of κ_1^e and c_1^0. If one is interested in the concentration of a heavy constituent instead of a light constituent, one need only replace c_1, c_1^0, c_1^1, and κ_1^e by c_2, c_2^0, c_2^1 and κ_2^e.

A numerical example given by these authors for diffusion of methane leads to a concentration of ^{13}C from 1·08 per cent to 19 per cent in an 8-metre tube at the rate of 12 mg. of ^{13}C per day. The experiments of Clusius and Dickel using larger temperature differences and smaller diameter tubes give about this predicted speed, so that it is at least of the right order of magnitude.

Wirtz [85] has considered two possible explanations of the thermal diffusion effect in liquids. The first of these assumes that the diffusing molecules are oscillating around equilibrium positions in the quasi-crystalline liquid. The diffusion coefficient would then be given by

$$D=f'\frac{\overline{l^2}}{6}2\nu e^{-Q/RT}, \qquad \ldots.(23)$$

where f' is a proportionality factor of order unity, l is the distance between two neighbouring rest positions, ν is the frequency of vibration, and Q is the activation energy per mole for the change in position. The thermal diffusion coefficient would then be defined as the temperature change of the diffusion coefficient, and, introducing numerical values, with Q approximately 4000 cal. for the sodium and chloride ions, one secures 10^{-2} for the ratio of D' to D. This is larger

than the experimentally estimated value of this quantity by a factor of 10, the experimental determination depending upon a Debye theory for the diffusion process, which as yet has not been published. The isotope effect would then depend upon a difference in the zero-point energies for the activation energy, but Wirtz concludes that this is too small an effect to account for the observed separations, since undoubtedly the vibration frequency must be very low, and hence the zero-point energy effect becomes negligible as in all cases of this kind.

The second suggestion is based upon a hydrodynamic theory of the process. In this case the diffusion coefficient is given by the equation

$$D=kT/6\pi r\eta.$$

Debye has suggested that the pressure which would be exerted upon a suspended body in the liquid would be proportional to the energy density, and the energy density in the case of thermal diffusion would vary as $c_v\partial T/\partial\chi$. On the one side of the ion or molecule there would be a higher pressure than on the other by this amount, and hence diffusion would occur. The force that would be exerted is given by the equation

$$K=f(4/3)\pi r^3 c_v \partial T/\partial\chi, \qquad \ldots.(24)$$

this being entirely analogous to the pressure exerted on a sphere suspended in a liquid in a uniform gravitation field, the force being equal to the volume multiplied by the gradient of the pressure. The velocity that the particle would acquire is given by the expression

$$v=\frac{K}{6\pi r\eta}=f(4/3)\pi r^3 c_v \frac{\partial T}{\partial\chi}\frac{1}{6\pi r\eta}, \qquad \ldots.(25)$$

and the ratio of D' to D is

$$\frac{D'}{D}=f\frac{4}{3}\frac{\pi r^3 c_v}{kT}. \qquad \ldots.(26)$$

This gives a result also of the order of magnitude of 10^{-2}.* This predicts that the thermal diffusion should increase as the cube of the radius of the diffusing particle, and thus accounts for the greater rapidity of diffusion of sulphate ions as compared with chloride ions.

Two difficulties come to mind with this theory. In the first place one must explain why it is that large macroscopic spheres are not rapidly propelled through liquids in which there is a thermal gradient. Also, one must explain why it is that each water molecule is not propelled through the water by the other molecules, i.e. why the pressure on the water molecules is zero. A macroscopic sphere would have a temperature gradient within itself which would exactly counterbalance the effect considered here; but this also means that large molecules cannot be expected to follow these formulae, since some sort of a temperature gradient would certainly exist in such molecules. It would appear, therefore, that if this theory is the correct one, it is due in some way to a different temperature gradient inside the diffusing molecule; and in the

* Wirtz has introduced a factor 12 in his equation, which the present writer cannot justify in any way and believes must be incorrect. This changes his ratio (which he reports as 10^{-3}) to 10^{-2}, as reported here.

case of like molecules the two gradients would be the same and hence diffusion would not occur. However, it is difficult to see how the two temperature gradients could be different in the case of isotopic ions or molecules except as it depends on the zero-point energies or similar factors, and this leaves the problem just as it was at the beginning, with the entire phenomenon to be explained. It would seem that the first suggestion of Wirtz is to be preferred.

6. *Centrifugal method of separation*

Lindemann and Aston [86] first suggested the centrifugal method for separating isotopes and Mulliken [87] first suggested an evaporative centrifugal method. Beams [88] and his co-workers developed this method using an air-driven centrifuge. The ratio of isotopes in the centre divided by their ratio at the periphery of the centrifuge, that is, the fractionation factor, is given by the equation

$$S = \exp \frac{(M_2 - M_1)\omega^2 r^2}{2RT}, \qquad \ldots.(27)$$

where ω is the angular velocity and r the radius of the centrifuge, T the temperature, and M_1 and M_2 the molecular weights of the constituents of the gas. It will be seen that apart from the temperature the only adjustable factor is the peripheral velocity of the centrifuge. Beams and Haines [88] have succeeded in getting a velocity of 8×10^4 cm. per second, and with the most recent chrome-molybdenum steels somewhat higher values could be secured in a small apparatus. With a peripheral velocity of 5×10^4 within the centrifuge and at 300° K., the fractionation factor should be about 1·05 for two isotopes differing in molecular weight by one unit, or 1·1 for a mass difference of two units. This factor depends only upon the difference of mass ; but of course the speed with which vapour can be withdrawn from such a centrifuge will depend upon the coefficient of diffusion, which will change with the atomic or molecular weight. Beams and Skarstrom [88] have secured a separation factor of 1·1 for the chlorine isotope in a small centrifuge and have produced slight changes in the proportions of the chlorine isotope.

Without some method of greatly increasing the effectiveness of this method it certainly cannot compete with other methods of separation. The author has suggested the use of a long rotating cylinder used as a fractionation column. A suggested design for such an apparatus would be approximately as follows : a relatively small tube, perhaps 0·5 cm. internal diameter and 3 metres long, would be rotated at high speed to produce a simple process fractionation factor of approximately 1·05 or perhaps 1·1. Liquid would be pumped into the top of this tube, would flow to the bottom, where it would be boiled by radiant heat. The vapour would flow upward mostly through the centre of the tube and escape at the top. Recently Goldwasser and Taylor [47] have described a very simple fractionation column of about these dimensions which has been useful in separating substances, the ratio of whose vapour pressures must be of the order of magnitude of the fractionation factor suggested above, and there seems to be no reason why such a centrifugal column should not be as effective as a stationary laboratory distillation column for the separation of organic liquids,

and only slightly more difficult to operate. With the suggested operation, heavy isotopes would concentrate at the bottom. In order to secure light isotopes it would be necessary to feed gas to the apparatus at the top, condense the vapour to a liquid at the bottom, and allow the liquid to escape at the top. As these centrifuges are constructed this could be done by boiling the liquid near the top in such a way that the incoming and outgoing vapours became thoroughly mixed.

7. *Mass-spectrograph methods*

The mass-spectrograph method of separating isotopes is so well known that no discussion of the fundamentals of the process need be mentioned here. It is only necessary to devise methods of collecting the deposits of the different isotopes on a glass or metal plate. This method has been used by Oliphant, Shire and Crowther [89] and by Rumbaugh [90] for the separation of the lithium isotopes. Limitations of the method are due to the small intensity of the atomic ions which can be secured for the ion beams. Smith and Scott [91] have described ion sources very much stronger than those previously used, and it seems that these will be sufficiently intense to make the mass-spectrograph method an important one for the separation of isotopes, particularly of the metallic elements.

§ 3. CONCLUSION

At the present time the most effective methods for the concentration of isotopes have been the chemical methods of electrolysis, distillation and chemical exchange reactions, the Hertz cascade diffusion method, and the thermal diffusion method of Clusius and Dickel. The last of these methods has proved within the last year to be a very interesting and effective method. It is limited to gaseous substances which are stable with the temperature differences which must be used, and which, at the temperature used, will not react with materials used for the construction of the apparatus. The Hertz diffusion method is also limited to gaseous substances. It appears that the thermal diffusion method is superior to the diffusion method because of the much simpler operation and the less expensive apparatus required. The rate of production in both cases appears to be of about the same order of magnitude, with perhaps an advantage in favour of the thermal diffusion method.

The chemical separation methods are limited to elements of low atomic weight and have been most useful in the separation of the hydrogen, carbon, nitrogen and oxygen isotopes, with the possibility that sulphur may also be separated by this method. In practice it will be difficult by these methods to produce entirely pure isotopes for the reason that the hold-up of the apparatus required is so large that long periods of operation are required to concentrate the working material in the apparatus. However, the rate of production by these methods is large as compared with any others so far devised. Deuterium is now an article of commerce and produced commercially in large quantities. Thode and Urey succeeded in producing about 0·75 gram of ^{15}N per 24-hour period, and there seems to be no reason why their apparatus cannot

be made to give many times this amount. The experiments in progress by Hutchison, Stewart and Urey are producing about 0·1 or more gram of ^{13}C per 24 hours, and it seems entirely possible that this can be increased by an obvious increase in size of their apparatus. The speed of the diffusion and thermal diffusion methods appears to be perhaps 5–10 mg. of ^{15}N or ^{13}C per 24-hour period, and there seems to be no obvious way to increase the capacity of the single apparatus much beyond this figure. Thus the chemical methods appear to be of the order of magnitude of 100 times as fast as the two diffusion methods in the case of these elements. The cost of operation and of the apparatus for the chemical methods is probably less than either of the diffusion methods. The cost of chemicals is a large item in operation, and runs to about 15 dollars per gram of ^{15}N or ^{13}C under laboratory conditions, while the cost of operation for the thermal diffusion method would appear to be many times these figures. Though development of the thermal diffusion method in the next few years may be rapid, it appears at present that the chemical methods will be most useful for the concentration of the isotopes of nitrogen, carbon, and oxygen, as well as hydrogen, and that the thermal diffusion method will be most useful for other elements which can be secured in gaseous compounds.

Two methods appear to be worth developing in addition to those mentioned above: the centrifugal fractionation column seems to the writer to give promise of being a very rapid method for the production of isotopes comparable to the production secured from small laboratory fractionation columns which are extensively used in the separation of organic liquids, and the progress that has been made on high-speed centrifuges indicates that the mechanical difficulties should not be too great.

The mass-spectrograph method appears as yet to be the only one adaptable to metallic elements in general, though some metallic elements have gaseous compounds, and hence the thermal diffusion method may be applicable to them. It seems to the writer that the high intensity of the ion sources developed by Smith and Scott should be adapted to mass-spectrograph use for the separation of isotopes and should make possible effective separation of the metallic elements.

With these methods available it seems that the separation of isotopes in adequate amounts for effective chemical and physical research is solved in principle and only requires persistent effort for actual production.

REFERENCES

(1) Soddy, F. *Trans. Chem. Soc.* **99**, 72 (1911).
(2) Aston, F. W. *Mass Spectra and Isotopes* (E. Arnold & Co., London, 1933).
(3) Harkins, W. D. and Madorsky, S. L. *Nature, Lond.*, **111**, 148 (1923); *J. Amer. Chem. Soc.* **45**, 591 (1923).
Harkins, W. D. and Mortimer, B. *Phil. Mag.* **6**, 601 (1928).
(4) Brönsted, J. N. and Hevesy, G. *Z. Phys. Chem.* **99**, 189 (1921); *Nature, Lond.*, **109**, 780 (1922); *Phil. Mag.* **43**, 31 (1922).
(5) Thode, H. G., Gorham, J. and Urey, H. C. *J. Chem. Phys.* **6**, 296 (1938).
(6) Washburn, E. W. and Urey, H. C. *Proc. Nat. Acad. Sci., Wash.*, **18**, 496 (1932).
(7) Lewis, G. N. and MacDonald, R. T. *J. Chem. Phys.* **1**, 341 (1933).

(8) ANDERSON, L. C., HALFORD, J. O. and BATES, J. R. *J. Chem. Phys.* **2**, 342 (1934).
BELL, R. P. and WOLFENDEN, J. H. *Nature, Lond.*, **133**, 25 (1934).
CHAMPETIER, C. *Bull. Soc. Chim. Fr.* [5] **2**, 162 (1935).
COLLIE, C. H. *Nature, Lond.*, **132**, 568 (1933).
ERLENMEYER, H. and GARTNER, H. *Helv. Chim. Acta*, **17**, 30, 1226 (1934); **18**, 419 (1935).
GOLDFINGER, P. and SCHEEPERS, L. *J. Chem. Phys.* **31**, 628 (1934).
HARKINS, W. D. and DOEDE, C. *J. Amer. Chem. Soc.* **55**, 4330 (1933).
HARTECK, P. *Proc. Phys. Soc.* **46**, 277 (1934).
HERTZ, G. *Naturwissenschaften*, **21**, 884 (1933).
MORITA, N. and TITANI, T. *Bull. Chem. Soc. Japan*, **10**, 257 (1935).
OTA, Y. *Mem. Fac. Sci. Agric. Taihoku*, **10**, 71 (1934).
SCHWARZ, K., KÜCHLER, L. and STEINER, H. *Z. Elektrochem.* **40**, 298 (1934).
TAYLOR, H. S., EYRING, H. and FROST, A. A. *J. Chem. Phys.* **1**, 823 (1933).
UREY, H. C. and WAHL, M. H. *Phys. Rev.* **45**, 566 (1935).
(9) ODDIE, T. H. *Proc. Phys. Soc.* **47**, 1060 (1935).
(10) TOPLEY, B. and EYRING, H. *J. Chem. Phys.* **2**, 217 (1934); *J. Amer. Chem. Soc.* **55**, 5058 (1933).
(11) BELL, R. P. and WOLFENDEN, J. H. *Nature, Lond.*, **133**, 25 (1934).
(12) COLLIE, C. H. *Nature, Lond.*, **132**, 568 (1933).
(13) BROWN, W. G. and DAGGETT, A. F. *J. Chem. Phys.* **3**, 216 (1935).
(14) EUCKEN, A. and BRATZLER, K. *Z. Phys. Chem.* A, **174**, 273 (1935).
(15) WASHBURN, E. W., SMITH, E. R. and FRANDSEN, M. *J. Chem. Phys.* **1**, 288 (1933).
(16) FARKAS, A. and FARKAS, L. *Proc. Roy. Soc.* A, **146**, 623 (1934).
(17) APPLEBY, M. P. and OGDEN, G. *J. Chem. Soc.* 163 (1936).
(18) WALTON, H. F. and WOLFENDEN, J. H. *J. Chem. Soc.* 1677 (1937).
(19) EUCKEN, A. and BRATZLER, K. *Z. Phys. Chem.* A, **174**, 269 (1935).
(20) HOLLECK, L. *Z. Angew. Chem.* **49**, 33 (1936); *Z. Elektrochem.* **44**, 111 (1938).
(21) HUTCHISON, CLYDE. Dissertation, The Ohio State University (1939).
(22) SELWOOD, P. W., TAYLOR, H. S., HIPPLE, J. A., Jr., and BLEAKNEY, W. *J. Amer. Chem. Soc.* **57**, 642 (1935).
(23) JOHNSTON, H. L. *J. Amer. Chem. Soc.* **57**, 484 (1935).
(24) HORIUTI, J. and OKAMOTO, G. *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **28**, 231 (1936).
(25) WALTON, H. F. and WOLFENDEN, J. H. *Trans. Faraday Soc.* **34**, 436 (1938). See this paper for general review of recent work.
(26) OKAMOTO, G., HORIUTI, J. and HIROTA, K. *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **29**, 223 (1936).
(27) FARKAS, A. *Trans. Faraday Soc.* **33**, 552 (1937).
(28) GURNEY, R. W. *Proc. Roy. Soc.* A, **134**, 137 (1932).
BELL, R. P. *J. Chem. Phys.* **2**, 164 (1934).
(29) LINDEMANN, F. A. and ASTON, F. W. *Phil. Mag.* **37**, 523 (1919).
LINDEMANN, F. A. *Phil. Mag.* **38**, 173 (1919).
(30) KEESOM, W. H. and VAN DIJK, H. *Proc. K. Akad. Wet. Amst.* **34**, 42 (1931).
(31) KEESOM, W. H., VAN DIJK, H. and HAANTJES, J. *Proc. K. Akad. Wet. Amst.* **37**, 615 (1934); *Physica*, **1**, 1109 (1934);
KEESOM, W. H., and HAANTJES, J. *Physica*, **2**, 981 (1935).
(32) UREY, H. C., BRICKWEDDE, F. G. and MURPHY, G. M. *Phys. Rev.* **40**, 1 (1932).
(33) SCOTT, R. B., BRICKWEDDE, F. G., UREY, H. C. and WAHL, M. H. *J. Chem. Phys.* **2**, 463 (1934).
(34) KEESOM, W. H. and VAN DIJK, H. *Congr. Int. Refrig.* **7**, 103 (1936).
(35) SCOTT, R. B. and BRICKWEDDE, F. G. *Phys. Rev.* **48**, 483 (1935).

(36) For the theory of this method see UREY, H. C. and TEAL, G. K., *Rev. Mod. Phys.* **7**, 54 (1935). See also CLUSIUS, K. and BARTHOLOMÉ, E., *Z. phys. Chem.* B, **30**, 237 (1935).
(37) LEWIS, G. N. and CORNISH, R. E. *J. Amer. Chem. Soc.* **55**, 2616-17 (1933).
(38) WASHBURN, E. W., SMITH, E. R. and FRANDSEN, M. *Bur. Stand. J. Res., Wash.*, **11**, 453 (1933).
(39) HENRIQUES, H. J. and CORNISH, R. E. *J. Phys. Chem.* **37**, 397 (1933).
(40) WAHL, M. H., HUFFMAN, J. R. and HIPPLE, J. A., Jr. *J. Chem. Phys.* **3**, 434 (1935).
(41) THODE, H. G. *Phys. Rev.* **56**, 214 (1939).
(42) KLAR, G. and KRAUSS A. *Naturwissenschaften*, **22**, 119 (1934); SMITH, E. R. *J. Chem. Phys.* **2**, 298 (1934).
(43) STEDMAN, F. *Canad. J. Research*, **13** B, 114 (1935)
(44) PEGRAM, G. B., HUFFMAN, J. R. and UREY, H. C. *Phys. Rev.* **49**, 883 (1936); *J. Chem. Phys.* **4**, 623 (1936); *Industr. Engng. Chem.* **29**, 531 (1937).
(45) HERZFELD, K. F. and TELLER, E. *Phys. Rev.* **54**, 912 (1938).
(46) BRICKWEDDE, F. G. *Symposium Amer. Assoc. Adv. Sci., Indianapolis*, 1937.
(47) GOLDWASSER, S., and TAYLOR H. S. *J. Amer. Chem. Soc.* **61**, 1751 (1939).
(48) UREY, H. C. and RITTENBERG, D. *J. Chem. Phys.* **1**, 137 (1933); **2**, 106 (1934).
(49) UREY, H. C. and GREIFF, L. *J. Amer. Chem. Soc.* **57**, 321 (1935).
(50) MURPHY, G. M. *J. Chem. Phys.* **5**, 637 (1937).
(51) UREY, H. C. and ATEN, A. H. W., Jr. *Phys. Rev.* **50**, 575 (1936).
(52) UREY, H. C., HUFFMAN, J. R., THODE, H. G. and FOX, M. *J. Chem. Phys.* **5**, 856 (1937).
(53) THODE, H. G., GORHAM, J. E. and UREY, H. C. *J. Chem. Phys.* **6**, 296 (1938).
THODE, H. G. and UREY, H. C. *J. Chem. Phys.* **7**, 34 (1939).
(54) FENSKE, M. R., TONGBERG, C. O. and QUIGGLE, D. *Industr. Engng. Chem.* **26**, 1169 (1934).
(55) UREY, H. C., ATEN, A. H. W., Jr., and KESTON, A. S. *J. Chem. Phys.* **4**, 622 (1936).
(56) ROBERTS, I., THODE, H. G. and UREY, H. C. *J. Chem. Phys.* **7**, 137 (1939).
UREY, H. C., MILLS, G. A., ROBERTS, I., THODE, H. G. and HUFFMAN, J. R. *J. Chem. Phys.* **7**, 138 (1939).
(57) HUTCHISON, C. A., STEWART, D. W. and UREY, H. C. Unpublished results.
(58) LEWIS, G. N. and MACDONALD, R. T. *J. Amer. Chem. Soc.* **58**, 2519 (1936).
(59) TAYLOR, T. I. and UREY, H. C. *J. Chem. Phys.* **6**, 429 (1938).
(60) HERTZ, G. *Z. Phys.* **79**, 108, 700 (1932).
(61) HERTZ, G. *Z. Phys.* **79**, 108 (1932).
(62) HARMSEN, H. *Z. Phys.* **82**, 589 (1933).
(63) HARMSEN, H., HERTZ, G. and SCHUTZE, W. *Z. Phys.* **90**, 703 (1934).
(64) WOOLDRIDGE, D. E. and SMYTHE, W. R. *Phys. Rev.* **50**, 233 (1936).
(65) WOOLDRIDGE, D. E. and JENKINS, F. A. *Phys. Rev.* **49**, 704 (1936).
(66) HERTZ, G. *Z. Phys.* **23**, 433 (1922); *Proc. K. Akad. Wet. Amst.* **25**, 434 (1922).
(67) MACGILLAVRY, D., Jr. Thesis, Columbia University, 1933; *Rec. Trav. Chim. Pays-Bas*, **56**, 330 (1937); *Trans. Faraday Soc.* **33**, 433 (1937).
(68) HERTZ, G. *Z. Phys.* **91**, 810 (1934).
(69) BARWICH, H. *Z. Phys.* **100**, 166 (1936).
(70) SHERR, R. *J. Chem. Phys.* **6**, 251 (1938).
(71) BLUMENTHAL, E. *Phil. Mag.* (7), **27**, 341 (1939).
(72) KOPFERMANN, H. and KRUGER, H. *Z. Phys.* **105**, 389 (1937).
(73) CAPRON, P., DELFOSSE, J. M., HEMPTINNE, M. DE and TAYLOR, H. S. *J. Chem. Phys.* **6**, 656 (1938).
(74) KRUGER, H. *Z. Phys.* **111**, 467 (1938).

(75) Clusius, K. and Dickel, G. *Naturwissenschaften*, **26**, 546 (1938).
(76) Clusius, K. and Dickel, G. *Naturwissenschaften*, **27**, 148 & 487 (1939).
(77) Groth, W. *Naturwissenschaften*, **27**, 260 (1939).
(78) Bramley, A. and Brewer, A. K. *J. Chem. Phys.* **7**, 553 (1939).
(79) Korshing, H. and Wirtz, K. *Naturwissenschaften*, **27**, 110 (1939).
(80) Clusius, K. and Dickel, G. *Naturwissenschaften*, **27**, 148 (1939).
(81) Korshing, H. and Wirtz, K. *Naturwissenschaften*, **27**, 367 (1939).
(82) Waldmann, L. *Naturwissenschaften*, **27**, 230 (1939).
(83) Grinten, W. van der. *Naturwissenschaften*, **27**, 317 (1939).
(84) Furry, W. H., Jones, R. C., and Onsager, L. *Phys. Rev.* **55**, 1083 (1939).
(85) Wirtz, K. *Naturwissenschaften*, **27**, 369 (1939).
(86) Lindemann, F. A. and Aston, F. W. *Phil. Mag.* (6) **37**, 523 (1919).
(87) Mulliken, R. S. *J. Amer. Chem. Soc.* **44**, 1033, 1729 (1922); **45**, 1592 (1923).
(88) Beams, J. W. *Rev. Mod. Phys.* **10**, 245 (1938), and references there cited.
Beams, J. W. and Skarstrom, C. *Phys. Rev.* **56**, 266 (1939).
(89) Oliphant, M. L., Shire, E. S. and Crowther, B. M. *Proc. Roy. Soc.* A, **146**, 922 (1934).
(90) Rumbaugh, L. H. *Phys. Rev.* **49**, 882 (1936).
Rumbaugh, L. H. and Hafstad, L. R. *Phys. Rev.* **50**, 681 (1936).
(91) Smith, L. P. and Scott, G. W., Jr. *Phys. Rev.* **55**, 946 (1939).

THE MESON

By R. PEIERLS
The University, Birmingham

RECENT work on the new particle, which I shall call meson,* has two aspects, which are as yet rather loosely connected with each other. One is the evidence on the properties of the meson from cosmic-ray experiments; the other is the rôle of the meson in the theory of nuclear forces. The first was dealt with in great detail in last year's *Report* on Cosmic Rays [41],† and I shall here summarize the main features and discuss the work published since that *Report* was written. I shall then deal more fully with the meson theory of nuclear forces, and, finally, with the applications of the theory to collision problems.

§1. MESONS IN COSMIC RAYS

The new particle was introduced [2], [67] in order to account for the penetrating component of cosmic radiation. The particle must have electronic charge, since all fast cosmic-ray tracks in cloud-chamber photographs show very nearly the same ion density. It must lose much less energy by radiation than the electron and, since it has the same charge, must be considerably heavier. It must, however, be lighter than the proton, since one very rarely finds tracks in which it ionizes heavily (i.e. in which its kinetic energy is less than its "rest energy" μc^2, where μ is its mass and c the velocity of light). Direct estimates of its mass are obtainable from photographs in which a meson has been caught near the end of its range and in which the ion density along the track, the curvature in a magnetic field, and the range, or at least two of these quantities, can be measured. Table 1 is a list of such determinations. The estimates of the experimental error, which are in all cases considerable, have not been quoted, since they are linked up with points of experimental technique. For these we must refer to the original publications.

The uncertainty in the value of the mass is still considerable, and there is as yet no conclusive proof that all mesons have the same mass, though this is probable.

Considerable difficulties in the interpretation of cosmic-ray data were removed by the assumption, taken from Yukawa's theory [76], that the meson

* None of the properties of the new particle seems to give rise to more controversy than its name. After the names "U-particle", "$\varkappa$-particle", "heavy electron", "yukon", "barytron", "dynatron", "mesotron", "mesoton" and "meson", and, for the neutral particle, "neutretto", had been used by different authors, the choice seems now to lie between "mesotron" [50] and "meson" [13], [24], of which I shall adopt the latter in this *Report*.

† I am indebted to Dr. Heitler for a number of very helpful discussions on the subject of the present *Report*.

Table 1. *Mass-determinations*

Mass (in units of electron mass)	Reference	Method (i = ionization density, c = curvature, r = range)
240	(54)	c, r, i
110		c, r, i
160		c, i
<200		c, r
200	(23)	c, i
200		c, i
250		c, r, i
350		c, r
220		
190	(75)	c, r, i
160		
430		
170	(57)	c, r
39		
55		
100		
112	(48)	c, r
120		
170		
569		

is capable of a spontaneous radioactive decay. This suggests that the attenuation of a meson beam after traversing matter should not depend only on the amount of matter traversed, but also on the distance travelled [28], [14]. If the matter is spread out over a greater distance, and hence the mesons spend more time in transit, a greater fraction of them can decay on the way. Hence the apparent absorption coefficient will be relatively less in dense materials. Precisely this had been found in comparing the absorption curve of vertical and of inclined penetrating rays in air with that in lead and other materials[14], [15], [61], [25], [5], [43], [20], [9], [21], [22]. These experiments indicate that the lifetime of a meson, in a coordinate system in which it is at rest, is between 2 and 4×10^{-6} sec.

It is probable that one of the products of the disintegration of a meson is an electron, and these decay electrons ought to form an appreciable part of the soft component of cosmic rays at sea level, in particular in light materials, in which these decay electrons have a long range. A careful theoretical discussion of the distribution of these decay electrons by Euler and Heisenberg [28] and comparison with the observed transition curve of cosmic rays from one material to another, provide strong evidence in favour of the existence of these decay electrons.

There is, however, also evidence of another process by means of which mesons may disappear without producing electrons [53]. This may be the capture of a meson by a heavy particle, cf. § 5.

If mesons are unstable, it follows that they cannot be part of the primary cosmic radiation, since they would not live long enough to travel through interstellar (or even through interplanetary) distances. They must, therefore, be created as secondaries in the atmosphere [26]. This follows also from another consideration : From the known energy distribution of the mesons it is possible to estimate the deflexion which they would have undergone in the earth's magnetic field if they were primaries [28], and there ought to be a considerable latitude effect. Actually, however, the penetrating component shows only a very weak latitude effect, and the suggestion is, therefore, that the mesons are secondaries, produced in the atmosphere by particles which are either more energetic, or uncharged, and are therefore not deflected by the magnetic field. It would seem likely that these primary rays are electrons, which create mesons either directly, or through intermediary γ-rays, but the possibility of their being protons [1], neutrons or photons does not seem to be ruled out completely.

There is some direct evidence of the production of mesons by γ-rays at sea level [63].

Mesons are also found in showers. This is proved most clearly by a comparison between measurements of the angular spread of extended showers and the cascade theory [29], [8], [7]. There seemed to be additional evidence from the absorption of shower particles, but this now seems doubtful in view of Janóssy's experiments [42].

Most of these showers are produced by the soft component of the cosmic radiation, but recent experiments seem to prove that mesons can themselves also produce showers. Schmeiser [64], [65] and Bothe [17] claim to have established the existence of " hard " showers which are mainly produced by mesons and consist of a few energetic mesons whose tracks form very small angles with the incident meson. Their conclusions are, however, opposed by Nielsen, Morgan and Morgan [56], whose experiments lead to different results.

Evidence of the production of secondary mesons by mesons passing through matter is supplied by experiments of Maier-Leibnitz [48], in which tracks of slow mesons are observed in a cloud chamber which is shielded by lead. These are too frequent to be interpreted as the tracks of originally fast mesons that have been slowed down in the lead.

The fact that mesons can produce secondary slow mesons seems in good agreement with the anomaly in the energy-loss curve of cosmic-ray particles, as found by Blackett and Wilson [74]. This process becomes impossible when the energy of the primary mesons is less than the rest energy, and therefore, if an appreciable fraction of the energy lost by fast mesons is due to this process, there will be an anomaly in the energy-loss curve at energies of the order of 100 Mev.

One of the causes of the disappearance of the meson is its spontaneous decay, but there is also evidence of its disappearing without the production of a fast electron or other ionizing particle [53]. In this case the energy must be transferred to a non-ionizing particle, such as a photon, or possibly to a neutral meson. Further evidence in favour of the existence of the latter is also

being accumulated (cf. [3], [74], [68], [63], [35], but against this [55]) by observing, in counters or in the cloud chamber, particles emerging from layers of different materials, while verifying that no ionizing ray entered the material at the same time.

§ 2. THEORY OF FREE MESONS

The existence of the meson was first assumed on theoretical grounds by Yukawa [78]. His argument was roughly as follows: It is known [60], [11] that the constituents of a nucleus (protons and neutrons) exert on each other forces other than the ordinary Coulomb repulsion between the protons; these "nuclear forces" are known to have a range of about 2×10^{-13} cm. By this one means that for distances exceeding this range the force between two particles decreases very rapidly, at any rate more rapidly than the inverse square law. Unless one wants to introduce direct action at a distance, which would be incompatible with relativity, one must assume that these forces are transmitted from one particle to the other by means of some field, in the same way in which the electromagnetic interaction between two charged particles is transmitted by the electromagnetic field. One knows that in the latter case the Coulomb law is a direct consequence of the wave equation

$$\nabla^2\phi-\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}=0 \qquad \ldots.(1)$$

for the electric-field quantities, which for a steady field reduces to

$$\nabla^2\phi=0, \qquad \ldots.(1\,a)$$

with the centrally-symmetric solution

$$\phi=\text{const}/r, \qquad \ldots.(1\,b)$$

valid for a point charge at the origin.

Clearly, if one wants to obtain a different law for the variation of the interaction potential with the distance, one has to assume that the equation for the nuclear field of force is different from (1). On general grounds it is difficult to abandon the assumptions that the field equations are linear and do not contain derivatives of higher order than the second. This, together with the postulate of Lorentz invariance, allows only one possible generalization of (1): *

$$\nabla^2\phi-k^2\phi-\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2}=0, \qquad \ldots.(2)$$

where k is a constant.

For steady fields this reduces to

$$\nabla^2\phi-k^2\phi=0, \qquad \ldots.(2\,a)$$

with the centrally-symmetric solution

$$\phi=\text{const}\,\frac{e^{-kr}}{r}. \qquad \ldots.(2\,b)$$

This function is of the required character in that it has a "range" $1/k$, and

* The sign of the λ^2 terms in the corresponding equations (14) and (18) of ref. (41) ought to be reversed.

one may hope to account for the observed nuclear forces by means of such a field theory if one assumes the constant k to be of the order of $\frac{1}{2}.10^{13}$ cm.$^{-1}$

Equation (2) is the relativistic wave-equation for a particle of mass

$$\mu=\frac{\hbar}{c}k, \qquad \ldots.(3)$$

where $\hbar$ is $1/2\pi$ times Planck's constant. Indeed if we apply (2) to a plane sinusoidal wave of wave-length $2\pi\chi\!\!\!^{-}$ and frequency $\omega/2\pi$, it implies

$$\omega^2/c^2=k^2+1/\chi\!\!\!^{-2}, \qquad \ldots.(4)$$

which, by using de Broglie's relations between wave-length and momentum, and between frequency and energy,

$$p=\hbar/\chi\!\!\!^{-}, \quad E=\hbar\omega \qquad \ldots.(5)$$

reduces to $$E^2/c^2=p^2+\mu^2c^2,$$

which is the correct relativistic equation between energy and momentum of a particle of the mass (3). This mass is of the same order of magnitude as that of the mesons observed in cosmic rays, and it seems therefore natural to identify those mesons with the particles associated with the nuclear field.

Before we proceed to discuss the interaction of the meson field with the nuclear particles we must first investigate the field equation in free space, i.e. the motion of a free meson.

The properties of the meson field are not completely fixed by adopting (2) [44]. It may be that the meson field is a scalar field, i.e. that it is described by a single function ϕ, which satisfies (2). In this case the wave associated with a meson is completely fixed if, at any given moment, we know ϕ and $\partial\phi/\partial t$ throughout space; $\partial^2\phi/\partial t^2$, and hence the values of ϕ at all subsequent moments, will then follow from the equation. In order to facilitate comparison with the other possibilities to be discussed below, we may write this equation in the following symmetrical form: At any moment we are given ϕ and the four quantities

$$F_x=\frac{\partial\phi}{\partial x}, \quad F_y=\frac{\partial\phi}{\partial y}, \quad F_z=\frac{\partial\phi}{\partial z}, \quad F_t=-\frac{1}{c}\frac{\partial\phi}{\partial t}, \qquad \ldots.(6)$$

between which there are identities, viz.

$$\text{curl } \mathbf{F}=0, \qquad \ldots.(7)$$

and they must then satisfy the equation

$$\text{div } \mathbf{F}+\frac{1}{c}\frac{\partial F_t}{\partial t}-k^2\phi=0, \qquad \ldots.(8)$$

which is equivalent to (2).

This scheme, however, does not represent the only possibility. If we adopt it, the wave function of a meson, and hence its physical state, is completely determined by its wave vector and frequency (in other words the momentum and energy of the meson). This theory then implies the assumption that the meson has no internal degree of freedom, in particular no spin.

A general theorem of quantum theory states that an elementary particle

may have a spin which is an integer or half-integer multiple of $\hbar$. For our purpose half-integer values would have to be ruled out, because such particles are associated with wave functions which are not single-valued, but which change sign if the coordinate system is turned through 360°. Such functions cannot represent a quantity whose sign has a physical significance, such as a force.

More physically, the reason for this is that particles of half-integer spin can never be created or annihilated singly by any kind of interaction with a mechanical system, since the angular momentum of such a system can only change by a whole number of units. Nuclear forces might be transmitted by particles of half-integer spin if these were only absorbed or produced in pairs. Such a theory was attempted by Heisenberg [36], [72] before the theory of Yukawa, using electrons and neutrinos instead of the meson, and gave a wrong order of magnitude. If it were found that the meson had a half-integer spin one would have to return to a theory of this type.

If we assume integer spin, the simplest case next to zero spin is that of one unit. Then the wave function has four components, which form a four-vector:

$$\varphi(=\phi_x, \phi_y, \phi_z\,;\ \phi_4),$$

which satisfies the equations

$$\left.\begin{aligned} k^2\varphi &= \frac{1}{c}\frac{\partial \mathbf{F}}{\partial t} - \operatorname{curl} \mathbf{G}, \\ k^2\phi_4 &= -\operatorname{div} \mathbf{F}, \end{aligned}\right\} \qquad \ldots.(9)$$

where

$$\left.\begin{aligned} \mathbf{F} &= -\operatorname{grad} \phi_4 - \frac{1}{c}\frac{\partial \varphi}{\partial t}, \\ \mathbf{G} &= \operatorname{curl} \varphi. \end{aligned}\right\} \qquad \ldots.(10)$$

As a consequence of these equations all field quantities satisfy (2). This case is closely analogous to the case of the electromagnetic field; by putting $k=0$ and replacing φ, ϕ_4, $\mathbf{F}$, $\mathbf{G}$ by the vector and scalar potentials, the electric and magnetic intensity, respectively, one obtains Maxwell's equations.

The spin might even be more than one unit, but this possibility has not been very well explored yet; at any rate the field equations would then become considerably more complicated.

For a reason which will become apparent at a later stage the possibility has also been discussed whether the meson equations might be a combination of (6), (8) with (9), (10), in other words, whether there might be mesons both of spin zero and of spin 1 [44], [52].

All these equations are written in such a way as to be Lorentz-invariant. This, however, still leaves undetermined the behaviour of the field quantities under an inversion of the coordinate system (change from right-handed to left-handed axes). This is connected with the question whether, for example, the vectors $\mathbf{F}$ in (18) or φ in (9) are polar or axial vectors (pseudo-vectors). In the first case one speaks of the meson as having an "even" character,

in the second case of it being "odd". In this terminology the photon is an even particle.

Having decided the spin and character of the meson it remains to consider its charge. There is no doubt that the cosmic radiation contains both positively and negatively charged mesons, and there may also be uncharged ones. The positive, negative and neutral mesons give rise to quite distinct fields.

§ 3. INTERACTION TERMS

We return for a moment to the analogy with Maxwell's equations. In the presence of matter these read

$$\left.\begin{aligned} \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} - \operatorname{curl}\mathbf{H} &= -4\pi\mathbf{j}, \\ \operatorname{div}\mathbf{E} &= 4\pi\rho, \end{aligned}\right\} \quad \ldots.(11)$$

$$\left.\begin{aligned} \mathbf{E} &= -\operatorname{grad} V - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \\ \mathbf{H} &= \operatorname{curl}\mathbf{A}, \end{aligned}\right\} \quad \ldots.(12)$$

where $\mathbf{A}$, V, $\mathbf{E}$, $\mathbf{H}$ stand for the vector and scalar potential and the electric and magnetic intensity in the usual manner. These equations are now inhomogeneous, including the effect of the charge density ρ and the current density $\mathbf{j}$ which produce the field. It is important to realize that one would also obtain a mathematically consistent theory if one were to add interaction terms in (12) rather than in (11). This would be appropriate for particles which could be regarded as true magnetic or electric dipoles, but there is no evidence of the existence of such particles.*

One may also think of particles carrying a multipole of higher order, but they would give rise to rather more complicated equations.

In the same way, the production of a field by heavy particles has to be represented in the equation by the addition of terms in (6) or (9), or in (8) or (10), respectively. One must then write

$$\mathbf{F} = \operatorname{grad}\phi + f\mathbf{s}, \quad F_t = -\frac{1}{c}\frac{\partial \phi}{\partial t} + fs_t, \quad \ldots.(6\,a)$$

$$\operatorname{div}\mathbf{F} + \frac{1}{c}\frac{\partial F_t}{\partial t} - k^2\phi = gw, \quad \ldots.(8\,a)$$

or

$$\left.\begin{aligned} k^2\boldsymbol{\varphi} &= \frac{1}{c}\frac{\partial \mathbf{F}}{\partial t} - \operatorname{curl}\mathbf{G} + g\mathbf{s}, \\ k^2\phi_t &= -\operatorname{div}\mathbf{F} + gs_t, \end{aligned}\right\} \quad \ldots.(9\,a)$$

$$\left.\begin{aligned} \mathbf{F} &= -\operatorname{grad}\phi_t - \frac{1}{c}\frac{\partial \boldsymbol{\varphi}}{\partial t} + f\mathbf{u}, \\ \mathbf{G} &= \operatorname{curl}\boldsymbol{\varphi} + f\mathbf{v}, \end{aligned}\right\} \quad \ldots.(10\,a)$$

* The magnetic moment of the electron does not come in this category, since quantum theory shows it to be a consequence of the charge of the electron, and the field produced by this moment is included in the term with $\mathbf{j}$ in (11). The situation would, however, be different for particles whose magnetic moment was not related to their charge by the Bohr relation; e.g. for neutral particles having a magnetic moment, as in Bethe's neutrino equations (9 *a*).

where f, g are constants which measure the two types of interaction. By analogy with the case of the electromagnetic field one may call g the "meson charge" and f the "meson moment" of the heavy particles. The other quantities refer to the state of the heavy particles. w is a scalar, the vector $\mathbf{s}$ $(=s_x, s_y, s_z)$ and s_t form a four-vector, and the two vectors $\mathbf{u}$, $\mathbf{v}$ form together an antisymmetric tensor; i.e. they are related like an electric and magnetic moment, or like electric and magnetic field.

The symmetry character of the quantities on both sides of any one of these equations must be the same, and the quantities which we can insert for w, $\mathbf{s}$, ... depend therefore on our assumption as to the symmetry character of the meson. So, for example, the vector $\mathbf{s}$, s_t is a true (polar) vector if the meson is even, and an axial (pseudo) vector if it is odd.

The quantities w, $\mathbf{s}$, s_t, $\mathbf{u}$, $\mathbf{v}$ refer to the state of the heavy particle or, speaking more exactly, to the transition of the heavy particle between two different states of motion, in just the same way in which in the quantum theory of radiation [38] the term governing the production of electromagnetic field by an atom refers to a transition of the atom from one state to another.

Moreover, if the meson carries an electric charge, conservation of electric charge requires that, on emission of a positive meson (m^+) a proton (p) must be transformed into a neutron (n), and vice versa, according to the scheme :

$$p \rightleftharpoons n + m^+,$$
$$n \rightleftharpoons p + m^-. \qquad \ldots.(13)$$

Neutral mesons would be analogous to light quanta in the sense that their emission or absorption would merely correspond to a change in the state of motion of the particle, without changing its nature :

$$p \rightleftharpoons p + m^0,$$
$$n \rightleftharpoons n + m^0. \qquad \ldots.(14)$$

We see from the above discussion that the details of the meson theory depend on our assumptions as to

(*a*) the spin of the meson,
(*b*) its symmetry character,
(*c*) the "meson charge" g and the "meson moment" f of proton and neutron,
(*d*) the rôles of charged and uncharged mesons.

In discussing the consequences of the meson theory for the nuclear forces we shall start with the simplest assumptions, viz. that the meson is even, has no spin, that there are only charged mesons, and that $f=0$. This is the theory adopted originally by Yukawa and his collaborators [76], [77].

In this theory we obtain an interaction between proton and neutron owing to the interaction of the one with the meson field of the other. Indeed, the proton has a tendency to become a neutron, with the emission of a positive meson. This process cannot actually take place in the sense that the meson can be set free, because this would require at least the energy μc^2, which is not available. However, as long as the meson is in the immediate neighbourhood

of the heavy particle, the interaction energy associated with the process of emission and re-absorption may make up the balance. Hence an isolated proton carries with it a field of mesons which are continuously being absorbed and re-emitted, but which cannot escape. This situation, which is familiar in quantum mechanics, is customarily described as the "virtual" existence of a meson.

Now if this meson, instead of being re-absorbed by the same heavy particle, is absorbed by a neutron which happens to be near, it will turn it into a proton, whereas the original proton is now left behind as a neutron. The net result is therefore that proton and neutron have changed places, and this interaction is therefore an exchange interaction of the type postulated by Heisenberg and Majorana [60] in order to account for the saturation character of nuclear forces. The interaction potential turns out to be of the form $(2\ b)$, the constant being equal to g^2.*

In order to account for the observed order of magnitude of the nuclear forces, the constant g would have to be about 5 electronic charges [34]. This has the effect that the dimensionless quantity $g^2/\hbar c$ is not much smaller than unity (whereas the corresponding quantity in the theory of radiation, $e^2/\hbar c$, is about 1/137). Now this quantity is a measure of the strength of the interaction, and the methods of solving the field equations, which are taken from radiation theory, are approximate methods based on the assumption that $g^2/\hbar c$ is small compared to unity. Hence the methods of approximation are not actually applicable, or, at any rate, it is not sufficient to consider only the first-order terms.

If one considers the details of this form of the theory, one finds the following defects:

(i) The force has a wrong sign. It gives repulsion for the ground state of the deuteron, in which the spins of proton and neutron are parallel [44].

(ii) It gives rise to a "Heisenberg" [60] force [44] rather than to a mixture of "Majorana" and "Heisenberg" forces, since, with the transfer of the electric charge from proton to neutron, the spin direction is not transferred. (This follows from the fact that the term on the right-hand side of $(6\ a)$ is isotropic and has no reference to direction in space. Thus the intermediate state of the meson, while it is being transferred from the proton to the neutron, bears no trace of the spin direction of the proton.)

(iii) As a consequence of (ii), there is no interaction between spin and orbit, i.e. the neutron-proton force does not depend on the angle between the line joining the two particles and their spin direction. This is in contradiction to recent evidence on the existence of a quadripole moment of the deuteron [46], which proves that the two particles have a tendency to make this angle small. One must, however, keep in mind that an effect of this nature may also be produced by relativistic terms (i.e. terms depending on the

* Different authors use different units in which to measure the meson field and the interaction terms. Our choice of units is such as to make g of the dimensions of a charge. It is convenient to give f the same dimensions as g, defining the moment by means of the length $1/k$, so that $f=g$ would correspond to a moment of magnitude g/k.

velocity of the proton and the neutron) and that therefore this difficulty may resolve itself.

(iv) The theory predicts an infinite self-energy for the heavy particle [34], [12], i.e. an infinite interaction energy of a proton or neutron with its own meson field. This difficulty is closely analogous to a corresponding difficulty in the theory of the electromagnetic field; both will probably be overcome in the same way, and notwithstanding this difficulty the meson theory might be used to the same extent to which the quantum theory of the electromagnetic field is successfully used to-day [38]. The same trouble, however, results in all higher approximations (i.e. terms containing higher powers of $g^2/\hbar c$) diverging. Owing to the large value of $g^2/\hbar c$ this produces more serious uncertainty than in the electric case. Möller and Rosenfeld [52] and Stueckelberg [68], [69], [70], [71], have discussed the influence of higher approximations and methods of obtaining more rigorous solutions, but serious difficulties remain.

(v) The theory gives no force between like particles (a proton emits only positive mesons, which cannot be absorbed by, and hence do not interact with, a second proton). Such forces would arise in the second approximation which, in view of (iv) above, may not be negligible. It is, however, known to-day that the forces between like and unlike particles are very nearly equal [18], [19], and this would not appear plausible if they arose from mechanisms so widely different.

These difficulties can, at least partly, be overcome if the heavy particles are given a "meson moment" f, if one allows the meson to have a spin, and if one assumes the existence of neutral mesons.

The introduction of neutral mesons solves (v). One might expect at first sight that, since unlike particles can interact both by means of charged and uncharged mesons, whereas like particles can only exchange neutral mesons, unlike particles ought to interact more strongly than like ones on this view. However, the forces produced by the neutral and charged meson fields in the case of unlike particles are not additive, one being an exchange force, the other an "ordinary" force. Kemmer [45] has shown that the constants in the equations can be adjusted in such a way that the forces between any two heavy particles are independent of the nature of the particles. Kemmer's equations are written in a symmetrical form in which only one kind of neutral meson occurs, but a generalization is possible in which one introduces two kinds of neutral mesons which are in the same relation to each other as the positive and negative electron in Dirac's theory, or as the positive and negative meson.

The sign difficulty, (i), is removed if one assumes the meson to have a spin of one unit (eqs. (9 *a*), (10 *a*)). Then, at least in the first approximation, the forces become attractive [44], [77]. If the term with the "meson-moment" is included, it is possible to obtain a mixture of Heisenberg and Majorana forces as required. However, the Majorana force always occurs together with a new term which behaves for short distances like the interaction between two dipoles, that is it depends strongly on the positions of the two particles with respect to their spin direction, and it increases with $1/r^3$ as the distance r

between the particles becomes infinitely small. The first property is welcome, since it helps to solve the difficulty (iii), but the second results in an infinite binding energy [10].*

It may be argued that a singularity which arises only on very close approach of the nuclear particles is not too serious, since for very small dimensions the methods of the present quantum theory will probably fail anyhow, for reasons which will be given below. The $1/r^3$ term in the interaction formula is therefore likely to become incorrect for very small r, and one may hope that in a complete theory this difficulty would disappear. In the theory of the electron similar difficulties have been provisionally overcome by cutting off the interaction term at a certain radius and leaving out everything that refers to smaller distances. Attempts have been made [34], [10] to do the same in the meson theory. One should, however, realize that the cutting-off certainly does not represent a consistent theoretical treatment; in particular because it is incompatible with relativity. One may reasonably apply such methods in cases in which the required result is practically independent of the exact distance at which one cuts off, as in most problems of radiation theory. Here, however, the results depend essentially on the radius of cut-off, and on the precise manner in which the remainder at small distances is treated, and for this reason such attempts can at the best represent only a very crude qualitative picture.

One of these cut-off theories is of particular interest. Bethe [10] tried to take the view that it would be unsatisfactory to assume the heavy particles to have both a meson-charge and a meson-moment. Since $f=0$ would lead necessarily to the difficulty (ii) he tried $g=0$. He showed that it is indeed possible to choose the cutting-off radius in such a way as to obtain the correct binding energies both for the triplet and for the singlet state of the deuteron. This theory gives the right order of magnitude for the quadripole moment of the deuteron if one assumes the interaction to be entirely due to neutral mesons, but not if one assumes Kemmer's symmetrical equations [45] with neutral and charged mesons.

The idea of charged mesons had been originally introduced in order to obtain exchange forces which would give saturation of the bonds in nuclei. Bethe's theory would give saturation by another mechanism, viz. owing to the strong angular dependence of the forces. Since a particle inside a large nucleus must necessarily have neighbours in all directions it is impossible to arrange them all in the most favourable directions. The details of this have not been worked out, but it is a possibility that has to be kept in mind.

One can avoid the difficulties with the $1/r^3$ term by assuming that there are both mesons of spin 1 and mesons with no spin, and arranging the

* In earlier publications on this subject it was often stated that this term would not contribute to the binding energy since its sign depends on the position of the particles and vanishes if all directions of the line joining the particles, relative to the spin direction, are equally probable. However, owing to the influence of the interaction term the particles will tend to arrange themselves along their spin direction rather more frequently than across it (as is proved by the existence of the quadripole moment), and hence the interaction energy due to this term becomes infinite.

respective meson-charges and moments in such a manner that the divergent terms just cancel [44], [52]. Such a theory may suffer from the difficulty (iii) since in the first approximation one does not obtain a dependence of the binding forces on direction, but such a dependence would arise (in analogy with the fine structure of atomic terms) if relativistic corrections were taken into account. Whether these corrections are sufficient to account for the observed quadripole moment remains to be seen.

§4. OTHER APPLICATIONS OF THE THEORY

Other results of the theory are essentially derived from the fact that a free proton may be regarded as being, for part of its time, transformed into a neutron plus a meson. So, for example, the magnetic moment of the proton as observed is actually an average over the two possible configurations. The fraction of time which it spends split into neutron and meson is small, but since the intrinsic magnetic moment of the meson is much greater than that of the proton owing to its smaller mass, even this small fraction of time contributes appreciably to the average moment. This may serve to explain the fact that the proton has a moment which is greater than the one it ought to have on the simplest form of Dirac's equation, and that the neutron has a negative magnetic moment although it has no charge.

Heitler and Fröhlich [31] have found that the values of the proton and neutron moment can be explained in this way if it is assumed that the fraction of the time which they spend in the transformed state is about $\frac{1}{3}$. However, if one tries to work out this fraction from the meson theory [34], one again finds divergent results, so that the details of this problem are beyond the range of the present theory.

In the same way in which the virtual existence of the meson gives rise to an anomalous magnetic moment, it also produces an anomalous charge distribution. While the proton is transformed into a neutron and meson, its electric charge is not localized exactly in the heavy particle, but spread over the meson field in the neighbourhood. This ought to give rise to small anomalies in the Coulomb field of the proton, and hence in the fine structure of the hydrogen lines. This effect has been discussed by Fröhlich, Heitler and Kahn [32], but here again the theory meets with serious divergences.

A similar effect arises in the interaction of a proton with a light wave, except that this does not depend only on the static distribution in space, but also on the motion during a transition of the proton into another state. Fröhlich, Heitler and Kahn [33] have shown that this has an important bearing on the nuclear photoeffect in the deuteron. Their results would predict that at γ-ray energies which are much greater than the threshold energy of about 2 Mev., the cross-section for photoeffect should be many times larger than that calculated on the assumption that the proton behaves like a point charge. No experiments in a suitable energy range seem so far to be available.

Yukawa added to his meson theory the assumption [79], [80] that the meson enters also as an intermediate particle into the mechanism of β decay, i.e. that in β decay one of the nuclear neutrons virtually splits into a proton and a

meson, and that the latter in its turn splits into an electron and a neutrino. This assumption is quite independent of the rest of the meson theory, and serves mainly for the unification of the theory.

If this assumption is made, it implies the existence of an interaction between electron and meson of a type rather similar to that between proton and meson. The strength of this interaction can be determined from the known constants of β decay.

This leads to the conclusion that the meson must be capable of disintegrating spontaneously into an electron and a neutrino. Indeed, if a transition is possible from a state involving the virtual existence of a meson to a free electron and neutrino, the same is necessarily true if we start with a free meson, and the same terms in the equations are responsible for both phenomena.

In order to give the correct relation between energy and life-time for β-radioactive nuclei, the electron has to be given a " meson charge " of about 10^{-7} electronic charges. With this, the life-time of a meson at rest comes out to be about $\frac{1}{2}.10^{-6}$ sec. The exact value is very sensitive to the mass of the meson that has been assumed, and to whether one gives the electron a " meson charge " or a " meson-moment " [78], [58]. This life-time is roughly of the same order of magnitude as the experimental figure quoted in § 1.

§ 5. COLLISION PROBLEMS

This section will deal briefly with the theory of processes involving free mesons. The simplest of these are the ionizing power and small-angle scattering of mesons, since they are mainly due to collisions in which the meson does not come very close to any of the constituents of the atom, and they therefore depend only on the electromagnetic interaction between the atom and the meson [74], [11a], and not on any short-range force acting on the meson.

The ionizing power depends only on the charge and mass, and the law of ionization has, in fact, been made use of in attributing one electronic charge to the meson. The scattering, which is independent of the mass (for a given momentum) is also normal, according to experiments by Blackett and Wilson [16]. The radiation is known to vary as m^{-2} with increasing mass of the particle. The detailed formula depends somewhat on the spin, owing to magnetic terms not being negligible, but at any rate the emission of radiation is much less frequent than for an electron and does not contribute appreciably to the average energy loss. Lastly, a meson may collide with an atomic electron and knock it out of the atom. The probability of this process has been calculated by Bhabha [11a] for mesons of spin $\frac{1}{2}$, and by Massey and Corben (in a paper in course of publication), who have shown that the probability of the production of fast knock-on electrons is very much greater if the meson has a spin of one unit. Bhabha and Carmichael * point out that the two values can be distinguished experimentally if it is assumed that large bursts of ionization produced by fast mesons arise by the production of a fast knock-on

* Private communication.

electron in the first instance which subsequently starts a cascade. Preliminary experimental results on the frequency of bursts at large depths seem to prove that the meson spin must be 1 if one wishes to account for the bursts in this way.

All other processes, however, involve the specific nature of the meson and its interaction with the heavy particles. Of such processes, the simplest, which were already contained in the virtual transitions important for the nuclear forces, are the reactions (13) and (14) read from right to left. They would correspond to the absorption of a meson. The excess energy and momentum might either be given up to some other constituent of the nucleus of which the heavy particle formed part, or given out in the form of radiation.

This process was considered in detail by a number of authors [39], [49], [62], [47]. It turns out that, for small meson energies, the cross-section for the process is inversely proportional to the velocity. Hence a slow meson has a constant probability, per unit time, of being absorbed by this process. In dense materials this probability is greater than that of spontaneous decay.

If the theory is carried to the second order of approximation, i.e. to terms involving the fourth power of the " meson charge " g, one obtains the possibility of a two-stage process, e.g.

$$m^+ + n \rightarrow p \rightarrow n + m^+, \qquad \ldots.(15)$$

where the intermediate state is only virtual, and the second meson has a direction different from the first. The result is, therefore, the scattering of mesons by nuclei. A detailed discussion of this process has been given by Heitler [39], [40]. It follows, then, that if one goes on to terms of yet higher order of approximation, one obtains multiple processes in which we may obtain, for example, instead of (15),

$$\left.\begin{array}{ll} & m^+ + n \rightarrow n + m^+ + m^- + m^+, \\ \text{or} & m^+ + n \rightarrow n + m^+ + m^0, \\ \text{or} & m^+ + n \rightarrow p + m^- + m^+, \end{array}\right\} \qquad \ldots.(16)$$

and similarly with still higher numbers of particles produced in one collision. Now the meson equations contain a constant k, which is of the dimension of an inverse length, and in processes involving mesons of a wave-length λ, the result depends on the dimensionless constant $k\lambda$. The general significance of this has been emphasized by Heisenberg [37]. In this case it turns out that the ratio of the probability of the multiple processes of higher order to that of the single process (15) is proportional to powers of $1/(k\lambda)$, so that for higher energies (small wave-lengths) the multiple process is more probable than the single one. This means, first of all, that our method of approach breaks down completely, since all attempts at solving the equations have so far been based on a method of successive approximations, which to the first order takes into account only the single process, in the second order a double one, and so on. If the multiple processes are more frequent than the single one, this method evidently does not work. Heisenberg [37] comes to the conclusion that the principles of quantum theory may no longer be applicable once this situation

arises. He develops a picture of what one may reasonably expect to obtain from a more consistent theory, and it would seem likely that the high probability of multiplication of fast particles in a single collision would remain. The question as to whether such elementary explosions are present to an appreciable extent in observable cosmic-ray phenomena is at present undecided.

Heitler [40] also discusses the probable breakdown of the theory in connexion with the divergences arising in the theory of nuclear forces (cf. § 3). If it is true that the methods of the present theory break down at high energies, this would also apply to the singularities of the theory which always arise from the contribution of virtual intermediate states of high energy. One may thus hope that a more consistent theory will result in changes which very roughly amount to the same as the present artifice of " cutting-off ", although the results of such a theory will, of course, be Lorentz-invariant.

Bhabha [13] points out that the fact that one obtains a large number of particles at one stroke need not necessarily imply the breakdown of quantum theory, since a similar situation arises in the theory of radiation in the region of very long waves, for which we have no reason to doubt the validity of the theory. It may be, therefore, that not the equations of quantum theory, but only our methods of solving them, are at fault. At any rate, there is no doubt that the calculations that have so far been made are not applicable if the wave-length of the meson is less than k^{-1}, i.e. if its energy is greater than its rest energy μc^2. Moreover these arguments also throw doubt on all other applications of the theory in so far as they are essentially influenced by terms referring to intermediate states of very high energies. It is not known at present what would follow from the equations of the meson theory if they were accurately solved, nor whether an accurate solution without singular terms exists.

No reliable theoretical formula is therefore available for any of the processes which involve a close collision of a meson with a heavy particle, except possibly for mesons of comparatively low energy.

Besides the processes already mentioned, it is of particular interest to consider (13), read from left to right, which may happen in two ways : Either a hard light-quantum hits a nucleus and gives part of its energy to a proton, which changes into a neutron, or *vice versa*, with the emission of a meson ; or a fast proton collides with another particle, and, owing to its being deflected, emits a meson, rather in the way that a fast electron can emit a light quantum when it hits an obstacle. Either of these processes may represent the origin of the mesons which are observed in cosmic rays. The first was discussed from this point of view by Nordheim [59] and by Massey and Corben [49] ; the second was suggested by Alfvén [1].

REFERENCES

(1) Alfvén. *Nature, Lond.*, **143**, 435 (1939).
(2) Anderson and Neddermeyer. *Phys. Rev.* **50**, 263 (1936).
(3) Arley and Heitler. *Nature, Lond.*, **142**, 158 (1938).

(4) Auger. *C.R. Acad. Sci., Paris*, **207**, 917 (1938).
(5) Auger, Ehrenfest, Fournier and Fréon. *C.R. Acad. Sci., Paris*, **204**, 257 (1937).
(6) Auger and Maze. *C.R. Acad. Sci., Paris*, **207**, 671 (1938).
(7) Auger, Maze, Ehrenfest and Fréon. *J. Phys. Radium*, **10**, 39, (1939).
(8) Auger, Maze and Robley. *C.R. Acad. Sci., Paris*, **208**, 1641 (1939).
(9) Barnothy and Forró. *Z. Phys.* **112**, 208 (1939).
(9 *a*) Bethe. *Proc. Camb. Phil. Soc.* **31**, 108 (1935).
(10) Bethe. *Phys. Rev.* **55**, 1261 (1939).
(11) Bethe and Bacher. *Rev. Mod. Phys.* **8**, 82 (1936).
(11 *a*) Bhabha. *Proc. Roy. Soc.* A, **164**, 257 (1938).
(12) Bhabha. *Proc. Roy. Soc.* A, **166**, 529 (1938).
(13) Bhabha. *Nature, Lond.*, **143**, 276 (1939).
(14) Blackett. *Nature, Lond.*, **142**, 992 (1938).
(15) Blackett. *Nature, Lond.*, **144**, 30 (1939).
(16) Blackett and Wilson. *Proc. Roy. Soc.* A, **165**, 209 (1938).
(17) Bothe. *Naturwissenschaften*, **27**, 305 (1939).
(18) Breit, Hoisington, Share and Thaxton. *Phys. Rev.* **55**, 1103 (1939).
(19) Breit, Thaxton and Esenbud. *Phys. Rev.* **55**, 1006 (1939).
(20) Bruins. *Proc. K. Akad. Wet. Amst.* **42**, 54 (1939).
(21) Clay. *Physica*, **6**, 114 (1939).
(22) Clay and van Gemert. *Physica*, **6**, 649 (1939).
(23) Corson and Brode. *Phys. Rev.* **53**, 773 (1938).
(24) Darwin. *Nature, Lond.*, **143**, 602 (1939).
(25) Ehmert. *Z. Phys.* **106**, 751 (1937).
(26) Ehmert. *Z. Phys.* **113**, 234 (1939).
(27) Ehrenfest. *C.R. Acad. Sci., Paris*, **206**, 428 (1938).
(28) Euler and Heisenberg. *Ergebnisse der exakten Naturwissenschaften*, **17**, 1 (1939).
(29) Euler and Wergeland. *Naturwissenschaften*, **27**, 484 (1939).
(30) Follett. *Proc. Phys. Soc.* **51**, 585 (1939).
(31) Fröhlich and Heitler. *Nature, Lond.*, **141**, 37 (1938).
(32) Fröhlich, Heitler and Kahn. *Proc. Roy. Soc.* A, **171**, 269 (1939).
(33) Fröhlich, Heitler and Kahn. *Proc. Roy. Soc.* A (in course of publication).
(34) Fröhlich, Heitler and Kemmer. *Proc. Roy. Soc.* A, **166**, 154 (1938).
(35) Froman and Stearns. *Phys. Rev.* **54**, 969 (1939).
(36) Heisenberg. *Zeeman Verhandelingen*. (The Hague, 1935.)
(37) Heisenberg. *Z. Phys.* **113**, 61 (1939).
(38) Heitler. *Quantum Theory of Radiation*. (Oxford, 1936.)
(39) Heitler. *Proc. Roy. Soc.* A, **166**, 529 (1938).
(40) Heitler. *Phys. Rev.* **54**, 873 (1938).
(41) Heitler. *Rep. Progr. Phys.* **5**, 361 (1939).
(42) Jánossy. *Proc. Camb. Phil. Soc.* **34**, 614 (1938).
(43) Johnson and Pomerantz. *Phys. Rev.* **55**, 104 (1939).
(44) Kemmer. *Proc. Roy. Soc.* A, **166**, 127 (1938).
(45) Kemmer. *Proc. Camb. Phil. Soc.* **34**, 354 (1938).
(46) Kellogg, Rabi, Ramsey and Zacharias. *Phys. Rev.* **55**, 318 (1939).
(47) Kobayashi and Okayama. *Proc. Phys.-Math. Soc. Japan*, **21**, 1 (1939).
(48) Maier-Leibnitz. *Z. Phys.* **112**, 569 (1939).
(49) Massey and Corben. *Proc. Camb. Phil. Soc.* **35**, 84 (1939).
(50) Millikan. *Phys. Rev.* **55**, 105 (1939).
(51) Möller. *Nature, Lond.*, **142**, 290 (1938).
(52) Möller and Rosenfeld. *Nature, Lond.*, **143**, 241 (1939).
(53) Montgomery, Ramsey, Cowie and Montgomery. *Phys. Rev.* **56**, 635 (1939).
(54) Neddermeyer and Anderson. *Phys. Rev.* **54**, 88 (1938).

(55) Nielsen. *J. Franklin Inst.* **226**, 601 (1938).
(56) Nielsen, Morgan and Morgan. *Phys. Rev.* **55**, 995 (1939).
(57) Nishina, Takeuchi and Ichimiya. *Phys. Rev.* **55**, 585 (1939).
(58) Nordheim. *Phys. Rev.* **55**, 506 (1939).
(59) Nordheim and Nordheim. *Phys. Rev.* **54**, 254 (1938).
(60) Peierls. *Rep. Progr. Phys.* **2**, 27 (1936).
(61) Rossi. *Nature, Lond.*, **142**, 993 (1938).
(62) Sakata and Tanikawa. *Proc. Phys.-Math. Soc. Japan*, **21**, 58 (1939).
(63) Schein and Wilson. *Phys. Rev.* **54**, 304 (1938).
(64) Schmeiser. *Z. Phys.* **110**, 443 (1938).
(65) Schmeiser. *Z. Phys.* **112**, 501 (1939).
(66) Shonka. *Phys. Rev.* **55**, 24 (1939).
(67) Street and Stevenson. *Phys. Rev.* **52**, 1003 (1937).
(68) Stueckelberg. *Helv. Phys. Acta*, **11**, 225 (1938); *ibid.* **11**, 299 (1938).
(69) Stueckelberg. *Helv. Phys. Acta*, **11**, 378 (1938).
(70) Stueckelberg. *Nature, Lond.*, **143**, 560 (1939).
(71) Stueckelberg and Patry. *Helv. Phys. Acta*, **12**, 300 (1939).
(72) Weizsäcker. *Z. Phys.* **102**, 572 (1936).
(73) Wentzel. *Phys. Rev.* **54**, 869 (1938).
(74) Williams. *Proc. Roy. Soc.* A, **169**, 531 (1938).
(75) Williams and Pickup. *Nature, Lond.*, **141**, 684 (1938).
(76) Yukawa. *Proc. Phys.-Math. Soc. Japan*, **17**, 48 (1935).
(77) Yukawa and Sakata. *Proc. Phys.-Math. Soc. Japan*, **19**, 1084 (1937).
(78) Yukawa and Sakata. *Nature, Lond.*, **143**, 761 (1939).
(79) Yukawa, Sakata, Kobayashi and Taketani. *Proc. Phys.-Math. Soc. Japan*, **20**, 993 (1937).
(80) Yukawa, Sakata and Taketani. *Proc. Phys.-Math. Soc. Japan*, **20**, 319 (1938).

THE PRODUCTION AND MEASUREMENT OF SHORT-WAVE RADIATIONS

By G. W. C. KAYE, O.B.E., M.A., D.Sc., F.R.S., G. E. BELL, M.Sc., Ph.D., W. BINKS, M.Sc. and W. E. PERRY, B.Sc.

Physics Department, The National Physical Laboratory, Teddington, Middlesex

§ 1. INTRODUCTION

Advances in the production of short-wave radiation and of fast particles are of particular interest to medical radiology. Thus the discovery of the neutron was immediately followed by investigations into the possible therapeutic value of the new particle. The main reason for the interest of the radiologist is that he hopes to find some radiation or particle which will have a more pronounced differential effect on normal and malignant tissue, and may therefore be of greater therapeutic value than radiations hitherto used. At the present time, there is much interest in the question of the influence of wave-length; whether, for example, the present tendency in the case of X radiation to use ever-increasing excitation voltages is warranted.

The present article deals with

(1) High-voltage apparatus for the production of X rays and high-energy particles;
(2) radium containers and measurements;
(3) the physical measurement of X-ray, γ-ray and neutron dosage; and
(4) precautionary measures to ensure adequate protection of X-ray and radium workers.

§ 2. HIGH-VOLTAGE APPARATUS

Apparatus for the production of high voltages may be divided into the following types:—

(*a*) Impulse generator.
(*b*) Cascade transformer.
(*c*) Condenser-rectifier voltage multiplier.
(*d*) Electrostatic generator.
(*e*) High-frequency resonance transformer.
(*f*) Low-frequency resonance transformer.
(*g*) Magnetic resonance accelerator (or "cyclotron").*
(*h*) Linear resonance accelerator.

* In accordance with the plan of this article, a description of the cyclotron, as of the other H.V. generators, is given in paragraph (*g*) below, although the cyclotron also forms the subject of a separate report by Dr. Mann (p. 125). [Editor.]

(a) *Impulse generator*

The principle of the impulse generator (or Marx circuit) consists in charging a number of condensers in parallel and then discharging them in series, whereby a momentary discharge occurs at a voltage equal to the sum of the voltages of the component condensers. Voltages up to about 3 Mv. have been obtained by this method. The main advantage is that the equipment is relatively cheap. When applied to x-ray tubes, the x-ray output is disappointing, owing to the short impulses (10^{-4} to 10^{-6} sec.) and to the relatively long rest periods (about 5 sec.). The circuit cannot be well applied to the hot-cathode x-ray tube, since an average tube current of 1 ma. would require a cathode emission of about 1000 amp. Further, the very high momentary energy values make severe demands on the target. Brasch and Lange [1], however, succeeded in obtaining a high emission from the cathode of a gas tube specially designed for operation with the impulse generator, and have produced x rays excited at voltages up to 2·4 Mv.

Joliot and his collaborators [2] have used a 3 Mv. impulse generator with a multi-sectioned tube of the Brasch and Lange type to liberate neutrons from beryllium by bombardment with the heavy bursts of x rays which occur at each impulse.

(b) *Cascade transformer*

The cascade-transformer method was first used for the production of high-voltage x rays in 1928 by Lauritsen and his collaborators. Four transformers, each giving about 250 kv., were arranged in series, the primaries of the second, third and fourth transformers being energized from low-voltage auxiliary transformers at the high-tension ends of the first, second and third transformers. The case of each transformer was insulated to a degree appropriate to its particular potential difference from earth. This generator gave an alternating-voltage output of about 1 Mv. (R.M.S.). In the first instance [3], it was applied to a continuously evacuated x-ray tube which worked at a maximum alternating potential of 750 kv., the anode being at earth potential. The electron current was obtained by " cold emission " from the electrodes. Owing to the erratic supply of electrons by cold emission, a hot filament was used in a later design [4]. For higher voltages it was found necessary to use a tube having both electrodes at high potential, with the central discharge chamber at earth potential. After " conditioning " for several weeks it was possible to operate the tube at 1100 kv.

The method has also been used to accelerate ions. Crane, Lauritsen and Soltan [5] obtained continuous currents up to 30 μa. of helium ions at voltages up to 950 kv. More recently, Crane [6] has obtained an ion intensity of 250 μa. at 1 Mv., using a five-sectioned accelerator and cascade transformer.

(c) *Condenser-rectifier voltage multiplier*

The condenser-rectifier voltage multiplier was developed independently by Greinacher, Schenkel, Villard, Cockcroft and Walton, and Bouwers. The principle of the method consists in arranging condensers and rectifying valves

in such a way that the voltage supplied by a main transformer can be multiplied several times and rectified.

Cockcroft and Walton [7] used the method in a high-voltage ion source which was the first to produce nuclear disintegrations. They used a circuit giving four-fold voltage multiplication. The rectifying valves enabled the condensers to be charged 60 times a second which, together with the smoothing effect of the condensers, resulted in an approximately constant potential. The discharge tube in which the ions were accelerated was in two sections for voltage distribution, since single sections could not be operated at voltages greater than 400 kv. A yield of 200 μa. of ions at 700 kv. has been obtained [8] by this method.

Allibone [9] has used the voltage-quadrupling circuit of Cockcroft and Walton in a million-volt x-ray unit installed at St. Bartholomew's Hospital, London. A perspective view of the installation is shown in figure 1.* The apparatus

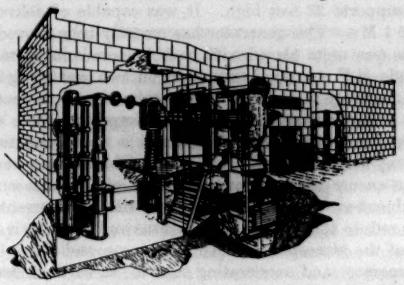


Figure 1. Perspective view of the million-volt x-ray installation at St. Bartholomew's Hospital, London.

is contained in three rooms, the treatment room, the cathode apparatus room and the anode apparatus room. The two latter rooms each contain a 600 kv. constant-potential generator, consisting of four rectifying valves and four condensers. The central part of the x-ray tube, which spans the treatment room, consists of a 14-inch diameter steel tube surrounded by a hollow steel drum which is filled with closely-packed lead shot for protective purposes. The drum is mounted on roller bearings at its ends, so that the emergent beam of x rays can be rotated about the tube axis. The floor of the treatment room can be raised and lowered, thus permitting the adjustment of the distance of the patient from the x-ray target.

In Bouwers' arrangement, a number of Villard units are cascaded. Two generators of this type [10], one for 1·25 Mv. and one for 2 Mv., have been supplied to the Cavendish Laboratory for nuclear research. Bouwers and Kuntke [11] have also constructed a 7-stage generator, producing 1·5 Mv. to earth, either positive or negative, according to the connections of the rectifiers.

Bouwers [10] has described a continuously evacuated neutron tube for

* We wish to thank Dr. T. E. Allibone for permission to reproduce this figure.

600 kv. Deuterons are produced and accelerated up to 300 kv. in one portion of the tube and further accelerated up to 600 kv. in the second portion. The two portions are separated by an earthed screen on which are mounted the vacuum pumps and gauges.

(d) *Electrostatic generator*

Van de Graaff [12] has produced voltages up to 5 Mv. by a simple electrostatic method, involving the charging of a well-insulated hollow conductor by swiftly moving belts, on which electric charges are sprayed at the lower end from a row of corona points at a potential of about 20 kv. The voltage attainable by this method is limited only by the breakdown of the air or by leakage along the insulators.

At Round Hill (Massachusetts Institute of Technology) a generator was built [13], consisting of two aluminium spheres 15 feet in diameter mounted on insulating supports 22 feet high. It was capable of delivering a current of 1·1 ma. at 5·1 Mv. This generator has recently been moved to Cambridge (Mass.) and the two units placed with their terminals in contact, thus giving in effect a single H.V. terminal supported on two insulating columns. The charging-belt assembly is installed in one column and the accelerating tube in the other. The target is located in an underground room and is at earth potential. For x-ray production [14] a target is used consisting of a water-cooled copper cup lined internally with lead. The electron beam is focussed into the cup to ensure the bombardment only of the lead surface. X rays have been produced at voltages up to 2·0 Mv., with tube currents of 8 to 16 μa.

In order to reduce the size of the apparatus and to make it independent of the humidity of the atmosphere, Herb, Parkinson and Kerst [15] enclosed an electrostatic generator and accelerating tube in an earthed steel tank 20 feet long and 5½ feet in diameter, containing compressed air at a pressure of 100 lb. per sq. in. At this pressure the breakdown strength of the air is nearly four times that of air at atmospheric pressure. A steady voltage of 2·16 Mv. to earth was obtained.

At the Carnegie Institution of Washington, D.C., a large generator is being constructed [16] in which the belts and accelerating tube are placed in a pear-shaped tank 55 feet high and 37½ feet in diameter. Compressed air at 50 lb. per sq. in. is to be used.

A million-volt x-ray generator of the electrostatic type is being used for cancer treatment and research at the Huntingdon Memorial Hospital, Boston [17]. The high-voltage terminal rests on an insulating column 10 feet high. Six four-ply fabric belts each 3 feet wide travel between the lower earthed pulleys and the corresponding pulleys 12 feet above, within the H.V. terminal. These belts run at a speed of about 4000 to 5000 feet a minute. The x-ray tube consists of twenty porcelain sections mounted vertically, the porcelain stack being 10 feet high. It is mounted on a steel flange supported from the floor. An earthed steel tube passes from this flange through the floor into the treatment room below, where it terminates in a water-cooled lead target. The generator operates normally at 1·0 Mv. with a tube current of 3 ma.

Further advances have been made as a result of utilizing the remarkable insulating properties of the gaseous form of dichloro-difluoro-methane (CCl_2F_2), known as "Freon". This gas has a breakdown strength about $2\frac{1}{2}$ times greater than that of nitrogen at all pressures up to 65 lb. per sq. in. and voltages up to 2 Mv. peak [18]. At pressures above atmospheric, Freon also shows a marked superiority in breakdown strength to transformer oil.

Trump and Van de Graaff [19] have recently constructed a 1·25 Mv. x-ray unit in which the electrostatic generator and accelerating tube are contained in a steel tank 100 inches high and 34 inches in diameter. With air at 165 lb. per sq. in., a maximum generator voltage of 1·4 Mv. can be obtained. With pure Freon at 50 lb. per sq. in., voltages up to 1·5 Mv. can be obtained. The use of Freon does not apparently produce any injurious effect on the materials used. The x-ray tube can be operated at all voltages from 200 kv. to 1·25 Mv. with a tube current of well over 1 ma.

(e) *High-frequency resonance transformer*

The Tesla transformer was the first form of high-frequency resonance transformer. Its modern counterpart, developed by Sloan [20], consists in passing high-frequency oscillations from a 200 kw. radio-oscillator through a single turn of copper tubing which forms the primary coil of a transformer. The secondary coil is a spiral of about 15 turns of copper tube, and the primary and secondary circuits are tuned to resonance. Consequently oscillations are induced in the secondary across which an alternating potential up to about 800 kv. is developed. One end of the secondary is earthed and the other carries the x-ray target. By placing a biassing voltage of about 10,000 volts on the cathode, the electron current only flows when the target potential is near its peak value. Since also the alternations are at a frequency of about a million per sec., the x radiation produced is almost the same as that from a tube excited by constant potential.

The main advantages of this generator are its freedom from exposed high voltages and its compactness, since it is possible to house the resonance transformer in a tank 4 feet high and 4 feet in diameter. The disadvantages are that the power consumption is very high and that cooling is difficult because of the great heat developed by eddy currents.

(f) *Low-frequency resonance transformer*

The most recent method [18] of producing high-voltage x rays is one consisting of a low-frequency resonance transformer within which the x-ray tube is mounted co-axially. Both are contained in an earthed steel tank less than 7 feet high and 4 feet in diameter, employing compressed Freon at 40 lb. per sq. in. as the insulation. The transformer has a primary coil of flat copper strip and a secondary coil of 120 thin flat sections, spaced apart for ventilation. The number of turns in the secondary is chosen to make its natural period of oscillation match that of the supply circuit, 180 cycles per second, a frequency obtained from a 60-cycle, three-phase power supply through the intermediary of a tripler unit. The x-ray tube is made up of 11 sections of glass tubing

sealed to metal rings, which carry the intermediate accelerating electrodes. Tappings are made at equal voltage intervals across the H.V. coil for connection to the various intermediate accelerating electrodes. The target is earthed and is mounted at the lower end of a chamber extending below the tank. X rays have been produced at voltages up to 1200 kv. with tube currents up to 9 ma. The unit is compact, there are no exposed parts at high potential and the target is readily accessible. As in the case of the Sloan oscillator, the power consumption is very high.

(g) *Magnetic resonance accelerator or "cyclotron"*

The methods of producing x rays and high-energy particles so far described might be called direct, that is, they are methods in which a potential is developed and applied to the x-ray tube or ion accelerator. Voltages up to 3 Mv., produced in air at atmospheric pressure, seem practicable by these methods, but the dimensions of the generators are large and consequently large buildings are required to house them. The voltages can be increased by placing the generators in high-pressure tanks but, even with this method, it is believed [21] that 6 to 8 Mv. will be the upper limit. Particles accelerated by voltages in excess of these have, however, already been realized by an indirect method (the "cyclotron"), due to Lawrence and Livingston [22], in which the particles are accelerated many times in succession by passing repeatedly through the same low-voltage field (probably not exceeding 100 kv.). The particles to be accelerated are produced in a vacuum chamber placed between the poles of an electromagnet. Within the chamber are two hollow, semicylindrical electrodes or D's, fed with high-frequency alternating current.

For particles having the same $\mathbf{e/m}$, the time of revolution in a circular path in the same magnetic field H is the same, irrespective of the radius of the path or of the velocity of the particle. If the electric field between the D's is adjusted to alternate with a period which is the same as the time of revolution of the particles, the latter will remain in "resonance" with the electric field. The particles will thus revolve in a series of increasing semicircles and will receive an increment of energy each time they cross the gap between the D's. The time of revolution of a proton is $6{\cdot}57\times10^{-4}/H$ sec. In a field $H=15{,}000$ oersteds, which is easily attainable, the time of revolution of a proton is $4{\cdot}38\times10^{-8}$ sec. The frequency of the alternating field must therefore be $2{\cdot}28\times10^{7}$ cycles per sec. As the time of revolution is proportional to $\mathbf{m/e}$ the frequency of the oscillations required to accelerate deuterons and α particles under the same conditions, will be $1{\cdot}14\times10^{7}$ cycles per sec.

The final energy acquired by the particles depends on the field strength and on the radius of the final path. In the Cambridge [21] and Liverpool [23] cyclotrons, the diameters of the magnet pole faces are 90 cm., permitting a D-radius of 38 cm., over which the magnetic field is sensibly uniform. With $H=15{,}000$ oersteds, the proton would reach a final speed of $5{\cdot}4\times10^{9}$ cm./sec., corresponding to an energy of 15 Mev. Up to the present time, outputs of 90 μa. of 8 Mev. deuterons and 1 μa. of 16 Mev. α particles have been reported [21]. The bombardment of beryllium with these high-energy deuterons

gives rise to a neutron yield of about 2×10^{12} per sec. Lawrence has recently constructed a cyclotron having a pole-face diameter of 145 cm., so that the geometrical energy limit is increased about $2\frac{1}{2}$ times. A difficulty arises, however, at these high energies. The speeds of the particles are becoming comparable with the velocity of light and there is a consequent relativistic increase in the mass of the particles. This causes an increase in the time of revolution within the D's. The particles when nearing the ends of their paths will therefore no longer be in resonance with the electric field and will consequently be lost from the beam. The particles could be made to arrive in phase with the electric field if their paths at large radii were decreased by a suitable increase in the magnetic field. Unfortunately such a magnetic field would have a defocussing effect upon the beam. Thomas [24] has suggested that, in addition to increasing the magnetic field with the radius, it should be made to vary periodically along the path of the particle, whereby focussing would be restored.

In April 1939 there were 26 cyclotrons, either in operation or under construction, in various parts of the world [25]. Seventeen of these are in the U.S.A. and three in this country.

(h) *Linear resonance accelerator*

The linear resonance accelerator is yet another application of the indirect method for producing high-energy particles. In this method, charged particles are directed along the axis of a system of tubular electrodes mounted co-axially. The first electrode is at earth potential, the second, fourth, sixth, etc., are connected to one side of a radio-frequency high-voltage source, while the third, fifth, seventh, etc., are connected to the other side. The voltage, the frequency of the oscillations and the lengths of the electrodes are adjusted so that a charged particle emerging from within the first electrode at the peak of the voltage cycle is accelerated into the second electrode, emerging from this half a cycle later to be further accelerated into the third electrode and so on through the system. Since the time interval between the accelerations is constant, the lengths of the electrodes must increase as the particle moves faster.

In one experimental arrangement [26], a high-frequency voltage of 42 kv. applied to 30 accelerating electrodes in line resulted in the production of a current of 0·1 μa. of 1·26 Mv. singly-charged mercury ions. The overall length of the 30 electrodes was about 114 cm.

In an improved apparatus [27], containing 36 electrodes to which a voltage of 79 kv. was applied, a mercury-ion beam of 0·01 μa. with energies of 2·85 Mv. was produced.

§ 3. RADIUM

The discovery of rich deposits of pitchblende in the North-West territories of Canada in 1930 has been followed by what is, for radium, a large-scale production, the output having already exceeded 100 g. The refinery at Port Hope, Ontario, has recently been enlarged and a monthly production of 9 g.

is now stated to be possible. The pitchblende resembles that found at Joachimsthal and the process of extraction is also similar, suitable modifications having been introduced to deal with valuable by-products, one of which is silver. After a preliminary roasting, the pitchblende is treated with salt in a reverberatory furnace in order to facilitate the subsequent removal of silver, and is then leached with water and sulphuric acid to which are added sodium nitrate, barium chloride and hydrochloric acid, the first to oxidize the uranium, the second to provide additional barium as a carrier for the radium, and the third to precipitate any silver which may still be in solution. The leaching is followed by filtration, the filtrate containing uranium, whilst radium, barium, silver, lead, silica and some calcium remain in the insoluble residue. Silver is removed by the use of sodium hyposulphite and sodium sulphide, and lead by treatment with caustic soda. The insoluble residue, containing silica, radium and barium, is then carbonated with soda ash and boiled in hydrochloric acid. On filtration, a residue of silica is left, the radium and barium being contained in the filtrate as chlorides. Agitation with sulphuric acid converts the radium and barium into sulphates which, after a number of purification processes, are converted into bromides and finally separated by fractional crystallization, the ultimate product being 90 per cent pure radium bromide. The complete preparation of the radium usually takes about two months and a ton of pitchblende (corresponding to nearly fifty tons of mined ore) yields on the average 100 mg. of radium, the extraction of which involves the use of about seven tons of chemicals. The percentage recovery of radium is about 95.

Units containing one or more grammes are employed in the treatment of cancer and other diseases by radium beam therapy. In these units the radium, in the form of highly concentrated sulphate, is enclosed in sealed metal containers, usually cylindrical tubes, each of which holds from 100 mg. to 250 mg. of radium element. Tubes and plaques for medical purposes other than beam therapy generally contain less than 100 mg., while needles for interstitial treatment contain amounts ranging from a fraction of a milligramme to several milligrammes. A mixture of radium and beryllium serves as a convenient neutron source for clinical and biological researches. Such sources usually consist of a hundred or more milligrammes of radium intimately mixed with a few grammes of finely powdered beryllium, the material being enclosed in a sealed metal container.

It is important that radium containers should be hermetically sealed to prevent leakage of radon and that their radium contents should not differ by more than a few per cent from those specified by the clinician responsible for their use. In the case of needles, plaques and certain types of tubes, uniform distribution of the radioactive material is also essential to permit accurate dosage. The tests to which containers are subjected at the National Physical Laboratory in order to ascertain whether or not they satisfy the above requirements have been described by Perry [28]. Leakage of radon is detected by allowing the container to remain in a sealed glass vessel with plugs of cotton wool for a few hours and then testing the cotton wool for

α-ray activity. Radium contents are determined by the γ-ray method, using an electroscope or a valve amplifier [29] as the measuring instrument. To examine the distribution of the radioactive material in the containers, contact photographs are obtained by placing the container in contact with a photographic plate for an appropriate period, during which the plate is exposed to a flash of light in order to register the outline of the container on the resulting negative.

The methods normally used for testing radium content do not indicate whether the γ rays are due to radium alone or to a mixture of radium, mesothorium and radiothorium. Tests for the presence of the latter substances are, however, sometimes required, e.g. in the case of a preparation extracted from a newly discovered source or of one the source of which is unknown. Methods based on the variation of the γ-ray activity of mesothorium with time and on the different absorption coefficients of the γ rays from the three radioactive substances have been used in such circumstances. These methods do not always, however, yield conclusive results, and for this reason Gorskov and Ionov [30] and Halban [31] have suggested that observations of the photonuclear effect produced by the γ rays from the test specimen in beryllium and heavy hydrogen might be used for purposes of quantitative differentiation. Further experimental investigation is, however, necessary before a definite opinion as to the advantages of this method over the older methods can be expressed.

During the last few years considerable attention has been given to the employment of counters for measuring weak α- and γ-ray activity. Mention has been made in previous *Reports* [32] of the investigations of Voigt, Evans and Finney in this connection. These authors' methods have been discussed by Poole [33], who points out that a combination of the emanation method with α-particle counting requires smaller quantities of test material than a direct counting method and reduces the effective influence of stray α particles on the ultimate sensitivity. The procedure used by Poole consists in liberating the radon from the specimen by fusion, transferring the radon to an ionization chamber and counting the α particles due to the radon and its succeeding products by means of a resistance-coupled amplifier operating a photographically-recording oscillograph. The apparatus records 3·6 particles per minute per 10^{-12} g. of radium, in agreement with the author's calculations [34], and permits radium-content determinations on a few grammes of rock.

Other purposes for which counting methods have been designed include the detection of lost radium containers and of stray radioactive deposits in radium laboratories and luminous-dial painting establishments, the detection of radioactive ores, and the measurement of the γ-ray activity of persons whose occupation exposes them to open radium preparations, e.g. those employed in radium refining, filling radium needles, etc. Evans and Mugele [35] have made an experimental study of the influence of the material and effective surface of the cathode, the pressure of the gas, and the operating voltage on the sensitivity of tube counters to γ rays. A portable unit [36] operated entirely by A.C. mains and incorporating a counting-rate meter is sufficiently sensitive

to detect 10^{-6} g. of radium one metre distant from the instrument. Units having this sensitivity and employing counters of similar design, but requiring a lower operating voltage (500–650 instead of 1200 volts) have been described by Curtiss [37]. The power supply for the first of these is derived from small dry batteries which, together with the associated high-speed counting equipment, are contained in a case, the total weight of the unit being 13 lb. The second unit is mains-operated and, with the exception of the tube counter, employs commercial radio components, the indicator being a milliammeter. The author states that the instrument retains its calibration to within 10 per cent over a period of several months and that only minor changes were required in order to adjust fifteen such instruments to the same sensitivity.

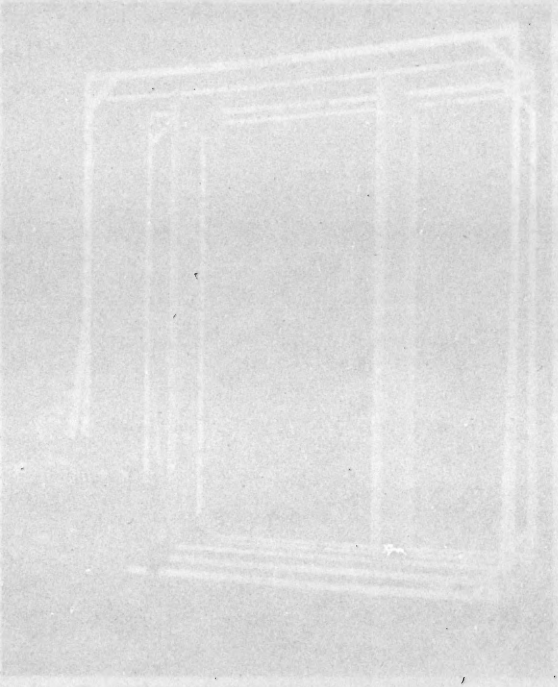
§ 4. X-RAY AND γ-RAY DOSIMETRY

In previous *Reports* [38], the development has been traced of a system for the quantity measurement of short-wave radiation as applied therapeutically. It has long been the accepted view that the effects of radiation in an irradiated medium can be correlated with the amount of corpuscular energy which is absorbed. From a physicist's point of view, therefore, a logical unit of quantity or dose of radiation is the energy absorbed per unit volume of the irradiated medium. Unfortunately, a direct measurement of the energy absorbed in tissue is not practicable and indirect means have been sought. Of these indirect methods, the ionization of air was recognized as the most suitable and led to the international adoption in 1928 of an ionometric unit of x radiation—the röntgen (r). The ionization of air is still accepted as the basis of radiation dosimetry, although doubts have arisen from time to time. In the present contribution we shall refer to some of the difficulties and to some of the many theoretical and experimental investigations which have been carried out in recent years in order to clarify the position.

Up to the time of the adoption of the röntgen, the maximum voltage used to excite the x radiation was about 200 kv. It was found that the ionization produced in a definite volume of air by such radiation could be measured by either of two types of chamber :—

I. *The "free-air" chamber*, usually of the parallel-plate type, in which a narrow beam of radiation limited in cross-section by a diaphragm system passes between parallel-plate electrodes. If (*a*) the guard plates are sufficiently large to ensure that the collecting electrode lies in an undistorted field and that, as a consequence, the ions which are collected are drawn from a length of path of the x-ray beam, equal to the length of the collector, (*b*) the separation of the electrodes is sufficient to ensure that the whole of the energy of the secondary electrons is given up in ionization before they reach the electrodes, and (*c*) the potential between the electrodes is sufficient to ensure saturation, then the chamber measures the full ionization produced by the secondary electrons within the "measured volume of air", which is not bounded by walls, but is defined by the diaphragm system and the length of the collector.

II. *The "air-wall" chamber*, in which the volume of ionized air is defined by a completely enclosing chamber (sometimes of "thimble" shape), the



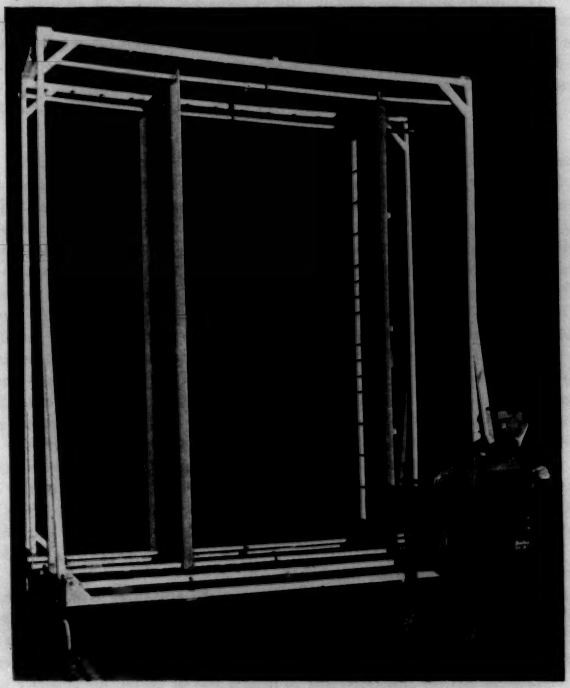


Figure 2. Large free-air parallel-plate ionization chamber.
(Reproduced by permission of the Council of the Royal Society)

[*To face p.* 105

walls of which are composed of a material having a mean effective atomic number the same as that of air.

In 1931 an intercomparison of the standard free-air chambers erected by the National Laboratories of this country, the United States of America, Germany and France, for the measurement of X rays excited at voltages up to 200 kv. showed that these agreed within the limits of error of observation (0·5 per cent). Up to this stage, the position of X-ray dosimetry was considered to be satisfactory.

Attempts to use the free-air chamber for the measurement in röntgens of supervoltage X rays and of the γ rays from radium failed, until it was realized that, owing to the long ranges of the electrons liberated in air by the passage of the radiation (amounting to about 2 metres in the case of radium γ rays), the condition laid down in the definition, that the secondary electronic energy should be fully utilized, necessitated the use of an exceedingly large parallel-plate chamber. Again, owing to the long ranges of the electrons, the ions which they liberate are no longer confined within the measured volume and there is loss of energy. On the other hand, some of the electrons liberated between the source and the measured volume now enter the latter, resulting in a compensation of energy. The definition of the röntgen was still satisfied if the incoming and outgoing energies exactly compensated, a condition which could only exist if the measured volume of air was everywhere bounded by a thickness of air in each direction equal to the maximum range of the electrons in that direction.

Kaye and Binks [39] constructed a chamber (see figure 2) having electrodes 3 metres high with a maximum separation of 3 metres and showed that, with such a chamber placed at least 2 metres from the radium source, the röntgen could be realized in the γ-ray region and ought therefore to be realizable for X rays of all wave-lengths. Nevertheless it was considered that the definition of the röntgen could with advantage be modified and a new definition was put forward, the wording following that of the old definition as closely as possible in order to preserve continuity. "The röntgen shall be that quantity or dose of X- or γ-radiation, whether direct or scattered, which, received at any point, would liberate under saturation conditions an ionic charge (of either sign) of one electrostatic unit per cubic centimetre in an indefinitely small volume of dry air at 0° C. and 76 cm. mercury pressure situated at that point, the air being bounded by an envelope of air or air-equivalent material of thickness equal to the maximum range of the secondary particles liberated by the radiation from the envelope and contributing to the ionization."

This unit permitted the measurement of dosage in röntgens, not only by the free-air chamber, but also by the air-wall chamber, provided the walls of the latter were of a thickness equal to the maximum electronic range in the material used (so called "thick" air-wall chamber).

Kaye and Binks stressed the limitations of the thick air-wall chamber and pointed out that a very thin air-wall chamber had a wider field of application, in that it measured the electronic energy absorbed at any point whether in free air, or on the surface of tissue or within it. It was realized, however,

that the thin-walled chambers were somewhat unsuitable for general therapeutic practice, partly because of their fragility and partly because of their susceptibility to the extraneous scattered electrons. Accordingly it was considered desirable to confine the modified definition of the röntgen to the thick-wall chamber. We shall, however, return later to the question of the need for supplementary measurements by means of the thin-walled chamber.

The behaviour of the air-wall chamber has been studied extensively [39]–[46] A general treatment by Gray [41] of the subject of dosage measurement by means of small air-filled chambers having walls of any material has done much to clarify the position. Gray's treatment may be summarized thus :—

(1) The introduction of a small air cavity into a solid or liquid medium does not disturb the number, speed and direction of the electrons crossing the surface which has become a wall of the cavity. This is only true if the dimensions of the air cavity are small compared with the range of the electrons.

(2) The cavity must be situated at a depth in the medium not less than the maximum range of the secondary electrons in the medium.

(3) The energy lost by the electrons in the air cavity (or the ionization produced in the air cavity) is related to the energy which they would have lost if they had traversed the same volume of the solid or liquid medium, in the ratio of the rates at which the individual secondary electrons lose energy in air and in the medium. If this ratio is ρ, then the energy, E_v, absorbed per unit volume of the medium is given by

$$E_v = \rho \epsilon J_v, \qquad \ldots.(1)$$

where ϵ=energy required to form a pair of ions in air, and J_v=ionization per unit volume produced in the air cavity.

(4) The ratio ρ for any medium can be expressed as

$$\frac{{}_eS_z}{{}_eS_{air}} \cdot \frac{n_z}{n_{air}},$$

where ${}_eS_z$ and ${}_eS_{air}$ are the stopping powers per electron of the medium (of atomic number Z) and of air, while n_z and n_{air} are the corresponding electron densities.

The relation between the energy absorbed per unit volume and the observed ionization may thus be written :—

$$E_v = \frac{{}_eS_z}{{}_eS_{air}} \cdot \frac{n_z}{n_{air}} \cdot \epsilon J_v. \qquad \ldots. (2)$$

(5) The electronic stopping power of a medium relative to air, ${}_eS_z/{}_eS_{air}$ (or f_z), is a slowly varying function of the atomic number of the medium and is practically independent of the speed of the electrons.

(6) It is assumed that ϵ has the same value of 33·0 ev. ($5{\cdot}28 \times 10^{-11}$ ergs) for all electrons having energies between a thousand and a million volts. Taking the electron density of air, n_{air}, at 0° c. and 760 mm. mercury pressure as $3{\cdot}92 \times 10^{20}$, expression (2) becomes

$$E_v = 1{\cdot}35 \times 10^{-31} . f_z . n_z . J_v \text{ ergs/c.c.} \qquad \ldots.(3)$$

(7) Gray concludes that the energy absorbed per unit volume (in other words, the dose) at any point in a medium may be determined absolutely, with an accuracy of about 5 per cent, by the measurement of the ionization in a small air-filled cavity introduced into the medium at the point under consideration.

(8) The equation $E_v = \rho \epsilon J_v$ implies that the ionization per unit volume of a cavity should be independent of the size of the cavity. But generally the equation is only true if, as assumed in (1), the dimensions of the cavity are small compared with the range of the electrons. For radium γ rays the maximum dimensions of the cavity should be of the order of a few mm., while for 200-kv. X rays, they should be of the order of 0·1 mm.

(9) If, however, the mean atomic number of the medium is adjusted to be equal to that of air, the secondary electron emission from the walls of the air cavity should be the same, apart from the density factor, as that from air and therefore the ionization per unit volume should be absolutely independent of the size of the cavity.

Applying these considerations to the air-wall chamber we see that, if the wall thickness is equal to the range of the secondary electrons, the chamber should give the true energy absorption per unit volume of air.

Gray was also led to propose a modified definition of the röntgen in the following terms :—

"The dose at any point is measured by the air-ionization equivalent of that part of the quantum energy which would be transformed into corpuscular energy by absorption in an infinitesimal volume of dry air at 0° C. and 760 mm. mercury pressure situated at that point, divided by the volume of the air, measured in cubic centimetres. If the amount of ionization is such that the total charge on the ions of one sign would be one electrostatic unit, then the dose is defined as one röntgen."

The physical basis of the definitions proposed by Gray and by Kaye and Binks is the same. It is possible to measure, in terms of these units, the electromagnetic radiation at a point in air or the dose in air. The units also afford an approximate solution to the problem of measuring the corpuscular energy absorbed per unit volume at any point in an irradiated medium such as soft tissue, since the mean atomic number of the latter is close to that of air. A dose measured in these units would not, however, be directly correlated with the actual energy absorbed in certain circumstances such as in the immediate vicinity of bone or skin, which contain elements having an appreciable photo-electric effect.

Another viewpoint is that expressed by Failla [42], [43] and his colleagues. They state that the "skin dose", an important factor in radiation therapy, cannot be measured by means of an air-wall chamber of the "thimble" type, since such a chamber is too large compared with the thickness of the skin. Both the air space and the solid walls of the chamber disturb the conditions obtaining at the surface of the skin, and therefore it is desirable to have a chamber "having no walls and no air volume". They have attempted to fulfil these conditions by means of an "extrapolation" chamber

(figure 3). In this, the air volume between two parallel-plate electrodes which form the chamber walls is first kept constant and the wall thickness gradually reduced to the practical minimum. A curve is then plotted showing the variation of ionization with wall thickness and a value for the ionization at zero wall thickness is obtained by extrapolation. Secondly the wall thickness is kept constant, at as low a value as is practicable, and the air volume reduced by varying the air thickness between the two electrodes. The ionization per unit volume of air is then plotted against air thickness, the curve being extrapolated to zero air thickness.

In later work [47] it was found that the ionization per unit volume was independent of the ionized volume up to 4 or 5 mm. separation of the electrodes, for 200-kv. x radiation with 0·5 mm. copper filter and for 120-kv. x radiation with no filter. Within this range of x-ray qualities, the extrapolation procedure seemed unnecessary. But with radium γ rays extrapolation was necessary, partly because of the geometry of the radium container and partly because of the

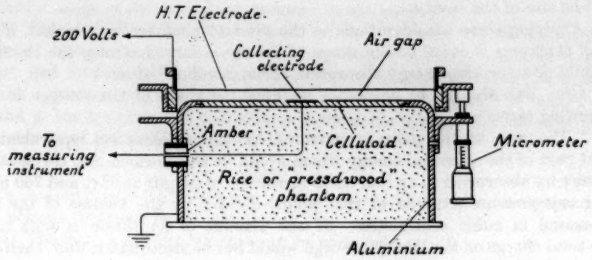


Figure 3. Extrapolation ionization chamber.

relatively short distances between the chamber and the radium, as a result of which a change in electrode separation altered the radium-chamber distance.

With the extrapolation chamber it is possible to determine the number of ions per c.c. (or per gramme) in air or at any point within an irradiated medium. Failla suggests that in this way a direct correlation between dose and effect becomes possible, and defines a unit of " tissue dose " as follows :—

" The tissue röntgen is the quantity of any ionizing radiation capable of producing $1{\cdot}62 \times 10^{12}$ ion pairs per gramme of air * at a given point in a given medium under the conditions in which the radiation is to be utilised."

Failla points out that the definition is general enough to cover any effect produced by ionizing radiation (including neutrons) irrespective of the nature of the effect, whether biological, chemical or physical. Accordingly the word " tissue " could be left out of the definition.

The criticisms levelled against the old (1928) definition of the röntgen

* That is, 1 e.s.u. of ionic charge per c.c. of air at 0° c. and 760 mm. mercury pressure.

resulted in the provisional adoption, at the Fifth International Congress of Radiology held at Chicago in 1937, of a modified definition as follows :—" The röntgen shall be the quantity of x or γ radiation such that the associated corpuscular emission per 0·001293 gramme of air produces, in air, ions carrying 1 electrostatic unit of quantity of electricity of either sign." (0·001293 gramme is the mass of 1 c.c. of dry atmospheric air at 0° c. and 76 cm. of mercury pressure.)

This definition could, in our opinion, be improved by the substitution of the phrase " is capable of producing " instead of the word " produces " The use of the latter word does not lead to a unique determination of the röntgen, as was intended. For example, it is possible for any number of röntgens to *produce* 1 e.s.u. of quantity of electricity if the corpuscular radiation is intercepted before the whole of its energy is expended in ionization. The suggested phrase would avoid any such ambiguity.

Another possible improvement would be to state that the 0·001293 gramme of air is " situated at the point at which the radiation is to be measured ". Otherwise the associated corpuscular emission could be liberated per 0·001293 gramme of air situated at any point A and could produce in air of no specified volume situated at another point B, at which the radiation is to be measured, ions carrying 1 e.s.u. of quantity of electricity of either sign. Such an interpretation must be avoided, since it would lead to dosage values depending on the relative positions of A and B.

The improvements suggested would make the new definition, as was the intention, fundamentally the same as those proposed by Gray and by Kaye and Binks, and consequently it suffers from the same imperfections as these units. It is doubtful, however, whether any single unit will suffice in all circumstances. The realization of this fact is leading at present to attempts to find a supplementary unit to the röntgen, since it is becoming clear that measurements of doses in röntgens do not provide the clinician with all the data he desires. A unit similar to the unit of " tissue dose " formulated by Failla might well fill the rôle. Holthusen [48] prefers to think of " electron dose " rather than tissue dose and maintains that the biological and therapeutic action of x or γ rays is determined ultimately by the local electron dose, that is, by the number of electrons liberated per unit volume at the point of irradiation. In his view " the local electron density is measured by the limiting value of the ionization per unit volume which is attained in an infinitely small chamber with infinitely thin walls ".

Assuming, therefore, that it is desirable to make measurements in terms of a supplementary unit of dosage, which will afford a closer approximation to the true energy absorption in tissue than does the röntgen, we are confronted with the problem as to how close this approximation should be. In making measurements with extrapolation chambers or very thin-walled " thimble " chambers, having walls of " tissue-equivalent " material, should air be used for convenience as the ionizing medium or should a gaseous mixture [49] be used having the same atomic composition as the tissue in question ?

Comparison of ionometric and calorimetric methods of energy measurement

By means of equation (3), the total energy absorbed at any point in an irradiated medium can be calculated from a measurement of the ionization produced in a small air-filled cavity introduced at that point. From measurements of the ionization produced at different points in a medium surrounding the source, it is possible to calculate the total energy emitted by the source. In order to determine the energy of the γ radiation emitted from Ra (B+C) in equilibrium with 1 gramme of radium, Gray [50] placed a radon source at the centre of a large block of aluminium, and measured the ionization at various points within the block. After allowing for the ratio of the electronic densities of aluminium and air and for the ratio of the electronic stopping powers of aluminium and air, it was estimated that the number of ion pairs which would be produced per second in air by complete absorption of the γ radiation from Ra (B+C) in equilibrium with 1 gm. Ra is $1{\cdot}90 \times 10^{15}$. This result may also be expressed in the form 8·7 calories per hour per gramme of radium.

It is interesting to compare this value with that determined by Zlotowski [51] from calorimetric measurements of the heating due to the γ radiation from Ra (B+C). The radioactive material was surrounded by a water jacket, the whole being enclosed in massive lead screens, which absorbed practically all the γ radiation. The heat developed by the α and β radiations was carried away by the stream of water flowing through the jacket. Zlotowski estimated the heating effect due to the γ radiation alone to be $9{\cdot}1 \pm 0{\cdot}15$ calories per hour per gramme of radium. This is in satisfactory agreement with the value determined ionometrically.

§ 5. NEUTRON DOSIMETRY

The absorption of fast neutrons in tissue is largely due to hydrogen atoms. The neutrons undergo elastic collisions with the hydrogen nuclei (protons) which recoil along short but highly ionizing paths. According to Chadwick [52] the recoil protons produced in tissue by the passage of 2-Mev. neutrons have an energy of about 0·3 Mev. and produce about 10,000 ions in a distance of some 4 to 5 μ. The ionization per unit path is about 100 times that produced by an electron ejected by hard x rays or γ rays. It is to be expected, therefore, that the passage of fast neutrons through tissue will produce a considerable biological effect and that, because of the higher ionization density along the path of the recoil protons, the biological effect may be different from that of x or γ radiation.

A number of attempts [53]–[59] have been made to compare the biological effects of neutrons and of x and γ radiation. This raises the problem of the measurement of neutron doses. Early attempts were confined to measuring the neutron dose by means of air-wall "thimble" chambers, using either air or methane as the ionizing medium. Lawrence and Lawrence [53] concluded from their first experiments that, when expressed in terms of the number of röntgens of x rays which produce the same ionization in the chamber, neutrons are roughly ten times as active biologically as x radiation (900 kv.) in respect of the alteration in the blood picture of the rat. Later work [54]

showed that the ratio of the number of röntgens of x rays to the number of "röntgens of neutrons" varied from about 2 to 5 according to the biological medium chosen. Aebersold [55] found that the intensity as measured by a bakelite thimble chamber at 15 cm. distance from a beryllium target bombarded by 20 μa. of 5·8-Mev. deuterons was the same as that due to 10 r/min. of x rays. This neutron intensity was found to be biologically equivalent to 20 to 50 r/min. of x rays, depending on the organism.

In this pioneer work, we see attempts to measure neutrons in terms of "equivalent röntgens" of x radiation. Zimmer [58], following the line of argument used by Gray in the case of γ radiation, shows that it should be possible to measure neutrons in röntgens. The energy per unit volume E_A produced per sec. in a small air-filled cavity in a chamber having walls of "Aerion" ($C_6H_{4\cdot4}O_{1\cdot6}$) and irradiated by fast neutrons is given by

$$E_A = \rho\epsilon J_{air}, \qquad \ldots.(4)$$

where J_{air} is the number of ion pairs per c.c. of air per sec. and ρ and ϵ have the meanings previously assigned to them. Further,

$$\rho = \frac{J_A}{J_{air}} = \frac{d_A \cdot S_A}{M_A} \cdot \frac{M_{air}}{d_{air} \cdot S_{air}}, \qquad \ldots.(5)$$

where d_A, d_{air} are the densities of "Aerion" and air,
S_A, S_{air} are the molecular stopping powers,
M_A, M_{air} are the molecular weights,
and J_A is the number of ion pairs per c.c. of "Aerion" per sec.

If n is the number of neutrons per second emitted by the source, the energy per c.c. of the "Aerion" chamber at a distance r from the source is given by

$$E_A = \frac{n}{4\pi r^2} \cdot \alpha \cdot E_n \cdot H_A \cdot d_A \cdot N\sigma, \qquad \ldots.(6)$$

where E_n = mean energy of the neutrons,
α = fraction of E_n given to the protons ($=1-1/e$),
H_A = fraction by weight of hydrogen in "Aerion,"
N = Avogadro's number,
σ = effective cross-section of protons for neutrons of energy E_n.

From expressions (4), (5) and (6) we get

$$J_{air} = \frac{n \cdot \alpha E_n \cdot H_A \cdot N\sigma M_A \cdot d_{air} \cdot S_{air}}{4\pi r^2 \cdot \epsilon \cdot M_{air} S_A} \qquad \ldots.(7)$$

For (Li+D) neutrons of mean energy 3·9 Mev.,

$$S_A/S_{air} = 4\cdot1 \text{ and } \sigma = 1\cdot7 \times 10^{-24} \text{ cm}^2$$

Therefore $J_{air} = 0\cdot296\ n/r^2$ ion pairs per c.c. per sec.
$= 1\cdot42 \times 10^{-10}\ n/r^2$ röntgens per sec.

Thus from a measurement of the ionization in röntgens per sec. at a point distant r cm. from a neutron source, it is possible to derive the number of neutrons per second emitted by the source. In this way, Zimmer estimated the number of neutrons per second to be $1\cdot0 \times 10^9$, which is in good agreement

with the value of $1{\cdot}2\times10^{9}$ which he obtained by means of a silver indicator. The agreement suggests that the ionometric method of measuring neutrons is fairly reliable.

Zimmer then proceeds to show that the ionization which 3·9-Mev. neutrons would produce in tissue would be about 1800 times greater than that which they produce in the "Aerion" chamber. On the other hand, for hard x rays the ionization in tissue would be about 1000 times greater than in the chamber. Thus if this type of neutron radiation produces 1 e.s.u./c.c. in an "Aerion" chamber, it should give the same energy absorption in tissue as 1·8 röntgens of hard x radiation.

In more recent work [59] it has been estimated that the mean dosage rate over the whole of a paraffin sphere of radius 5 cm., with a Ra+Be source of neutrons at its centre, is about 2×10^{-3} r per hour per mg. of radium. Comparing this with the γ-ray dosage rate of 8·3 r per hour at 1 cm. from 1 mg. of radium it is seen that the ionization due to neutrons from a Ra+Be source is less than 1 per cent of that due to the γ radiation.

Gray and Read [60] have calculated that neutrons of about 3 Mev. energy which produce 1 e.s.u. per c.c. in a graphite chamber, having an air volume in the form of a disk 2 cm. in diameter and 0·25 cm. deep, should result in the same energy absorption in tissue as 7 röntgens of γ rays. From an experimental comparison of their graphite chamber and of Zimmer's "Aerion" chamber, Gray and Read deduced that the ionization in the latter should be multiplied by a factor of 2·1 to obtain the energy absorption in tissue. The agreement with Zimmer's figure of 1·8 is satisfactory.

Even though the energy absorption in tissue is accepted as a reasonable basis of comparison of neutron doses and of x- and γ-ray doses, we are still confronted with the fact that the ratio of the neutron dose to the x- or γ-ray dose which produces the same biological effect seems to vary according to the biological medium chosen. Thus Zimmer and Timoféeff-Ressovsky [61] find that, in order to produce the same mutation rate of Drosophila eggs, the amount of energy absorbed from an x-ray beam excited at 100 kv. is only about 0·65 times the amount of energy absorbed from (Li+D) neutrons. They show, however, that x radiation produces the same effect as radium β rays or radium γ rays. On the other hand, Gray and Read [62] find, from a comparison of the lethal effect of fast neutrons and radium γ rays on the growing tips of broad-bean roots, that the ratio of γ-ray energy to neutron energy is about 11. This leads to the conclusion that the energy absorbed in tissue is not the only physical factor to be taken into account in comparing the biological effects of neutrons and of x and γ radiation.

§ 6. X-RAY AND RADIUM PROTECTION

When the discovery of x rays was announced by Röntgen in 1895 it was at once realized that they were likely to prove of service to medicine, and in a very short time the necessary apparatus, then consisting of an induction coil and an x-ray tube, was installed in a large number of hospitals. There are

numerous records of medical radiographs taken in 1896, and almost simultaneously the fluorescence of barium platino-cyanide and other materials was turned to use in the study of fractures and other simple lesions. The apparatus occupied little space, and as complete darkness was essential it is not surprising that the early x-ray departments were located in disused cellars, small storerooms and other unhygienic situations.

During the ensuing years, x rays continued to be used to a progressively increasing extent, until by 1912 it had come to be realized that persons who had worked for several years with x rays suffered injuries and deterioration in general health which could only be attributed to the special circumstances of their work. These injuries included superficial damage to the hands, often resulting in cancerous growths, and also damage to internal organs which gave rise *inter alia* to a form of anaemia. In the more progressive x-ray departments steps were taken to combat this by shielding the operators from as much radiation as possible, and by providing healthy surroundings, but the subject of protection was not systematically studied until the British X-ray and Radium Protection Committee was formed in 1921 as the result of co-operative action between the Royal Society of Medicine, the Röntgen Society, the British Association for the Advancement of Radiology and Physio-Therapy (now the British Institute of Radiology incorporated with the Röntgen Society), the Institute of Physics, the Radium Institute (London), and the National Physical Laboratory.

Similar committees were later set up in other countries, and in 1927 a standing International Protection Commission was constituted. The British Committee which met under the chairmanship of Sir Humphry Rolleston realized that dangers arising from x-ray work might be due to three main causes, namely, (1) the direct action of the radiation, (2) the effects of nitrous fumes produced in the atmosphere by the high-tension equipment used, and (3) the electrical risks associated with the operation, often in the dark, of apparatus working at many thousands of volts and having high-tension conductors which were inevitably somewhat exposed.

Protection from radiation

As regards protection from the direct action of the radiation, there proved to be little evidence in those days as to the maximum amount of radiation which could be tolerated indefinitely (the so-called tolerance dosage rate), and experiments to determine this were carried out in many hospitals both in this country and abroad where the working conditions appeared to be reasonably good and the workers had remained in normal health for some years. This work was complicated by the fact that at that time there was no satisfactory physical unit for the measurement of quantities of x radiation, and the results had to be expressed in terms of a somewhat uncertain biological unit, the "erythema" dose. Various attempts were made to express the erythema dose in terms of the röntgen and somewhat wide variations were reported, but for dosage rates commonly used a reasonable mean value appears to be about 550 röntgens. Mutscheller [63], making measurements in several typical good

departments, concluded that 1 per cent of the erythema dose per month was harmless (i.e. about 0·2 r per day), while Barclay and Cox [64] found a case in which exposure to 0·007 erythema dose per day (i.e. about 4 r per day) had produced no ill effects in six years, and concluded arbitrarily that 1/25 of this dose (i.e. 0·16 r per day) might be considered as a safe tolerance dose.

The basis of modern x-ray protection may be said to date from the general acceptance of a tolerance dosage-rate of 1/1000 of an erythema in 3 days, corresponding to 0·2 r per day, or 10^{-5} r per sec. for continuous irradiation during an 8-hour day. At the present time most x-ray departments have protection such that the dosage-rates do not exceed this value, and it is unusual to find that the health of the workers in the x-ray departments is inferior to that of other hospital staffs enjoying a similar standard of living. It has been claimed by some that the human body can tolerate indefinitely considerably more radiation than 0·2 röntgen per day, but as the amount of stray radiation can in the majority of cases be reduced to this value without impairing the functioning of apparatus there would appear to be no grounds for adopting a less stringent basis for protection. On the other hand, it has been claimed that certain organs are damaged by doses much less than the accepted tolerance dose.

Somewhat similar considerations appear in formulating recommendations for radium protection, except that in this case the actions of both β and γ rays and sometimes α rays have to be taken into account. The β rays are practically all absorbed in the skin and so give rise to superficial injuries, while the γ rays, being highly penetrating, affect the internal organs and tend to cause serious blood changes. There is every reason to expect that γ rays will behave like x rays, and it has generally been assumed that the tolerance dosage rate is of the same order,* although no direct measurements appear to have been attempted on human subjects. Recently Russ and Scott [65] have studied the effect of weak γ radiation applied to rats over a long period of time (168 days) and have found that dosage rates of 50 times the tolerance have very pronounced injurious effects. These experiments suggest that, at least so far as γ rays are concerned, the tolerance dose of 0·2 r per day is probably near to the limit of safety and thus provide further argument against relaxing the standard of protection adopted internationally.

The next step in formulating a scheme of protection is clearly that of determining what thickness of absorbing material is necessary to reduce the intensity of the x-ray or γ-ray beam at any point to the tolerance dosage rate. In this connection it is customary to express the amount of protective material required in terms of lead, since this is the most highly absorbing material available at reasonable cost and in suitable form for the purpose.

Figure 4 gives a series of curves showing the absorption of x rays from a tungsten target excited at various voltages up to 400 kv., while figure 5, derived from it, facilitates the determination of the thickness of lead required to reduce the dosage rate of any x-ray beam (up to 400 kv.) to the tolerance

* It is now well known that equal doses of x rays and γ rays produce substantially the same biological effects in the majority of circumstances (e.g. Crabtree & Gray, *Brit. J. Radiol.* **12**, p. 39 (1939)).

value. For example, if in the absence of any protective material the dosage rate at any point due to 200-kv. x rays is 3 röntgens per minute, the thickness of lead required for protection purposes is read off as 4 mm. The curves of

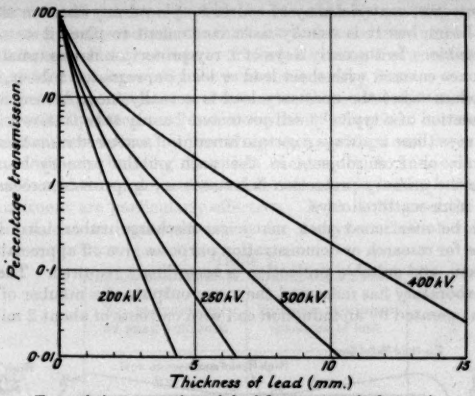


Figure 4. Transmission curves through lead for x rays excited at various voltages.

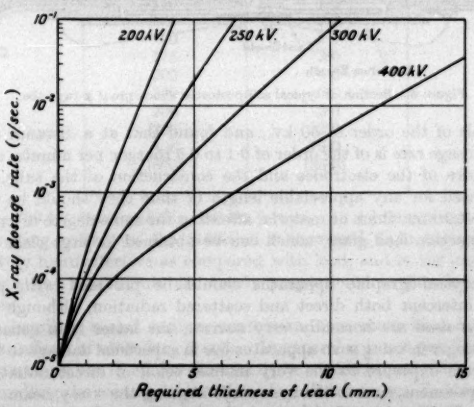


Figure 5. Lead protection curves for x rays excited at various voltages.

figure 5 relate, of course, to continuous irradiation during an 8-hour day: if the apparatus is used intermittently the amount of protection necessary may be reduced. For example, if the apparatus is used for only one hour per

day, it may be regarded for the purpose of reading the protection required from figure 5 as having one-eighth of the output, and the corresponding lead thickness required is only 3 mm.

The protective material may of course be placed anywhere in the path of the x-ray beam, but it is usually most convenient to place it as near to the tube as possible. In the early days of x-ray protection it was usual to enclose tubes in boxes covered with sheet lead or lead-impregnated rubber, but in the case of modern tubes the necessary lead is actually incorporated in the tube itself. A section of a typical " self-protected " x-ray tube is shown in figure 6. In using x rays there is always a certain amount of scattered radiation produced from the air or from objects in the path of the primary beam, and in addition to the primary protection it is necessary to provide screens to afford protection from scattered rays.

It may be mentioned that many gas-discharge tubes used in physics laboratories for research or demonstration purposes give off appreciable quantities of x rays, and suitable protection is accordingly required. The National Physical Laboratory has measured the x-ray outputs of a number of discharge tubes when operated by an induction coil with currents of about 2 milliamperes

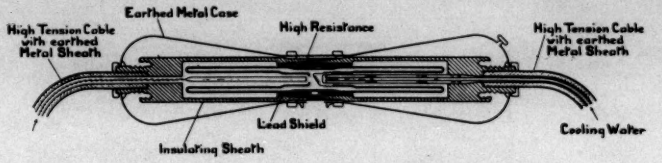


Figure 6. Section of typical self-protected shock-proof x-ray tube.

at potentials of the order of 50 kv., and found that at a distance of 50 cm. the x-ray dosage rate is of the order of 0·1 to 0·5 röntgen per minute, depending on the nature of the electrodes and the construction of the tube. If such tubes are used for any appreciable length of time they should be protected by lead sheet 1 mm. thick or material affording the same degree of protection ; in this connection lead glass, which can be obtained in large clear sheets, is very useful.

X-ray crystallographic apparatus should be provided with protecting screens to intercept both direct and scattered radiation, although since the x-ray beams used are normally very narrow, the latter is in general small. The chief danger in using such apparatus lies in superficial damage to the hands resulting from exposure to the very intense beam of direct radiation while adjusting specimens, and to the eyes while aligning the x-ray beam.

Table 1 [66], [67] gives the thicknesses of lead recommended for average conditions. Although primarily based on hospital practice, these thicknesses are applicable to the majority of conditions under which x rays are used for industrial and research purposes ; should screening be carried out continuously over long periods the protection should be increased somewhat.

In the case of radium γ rays the problem of protection again involves the absorbing of the rays by means of suitable screens, and table 2 [68] gives the thicknesses of lead required to afford protection at various distances from various amounts of radium. It will be seen that by working at relatively large distances from the radium, the thickness and hence the weight of absorbing screens can be appreciably reduced ; in fact, doubling the working distance is roughly equivalent to increasing the lead thickness by 3 cm.

Besides providing adequate protection for the actual operators, it is often necessary to afford protection to ensure the safety of workers in rooms adjacent to X-ray and radium rooms, by providing walls, ceilings and floors of suitable materials. Ordinary building materials such as brick and concrete afford a considerable measure of protection, while concretes including barium sulphate as one constituent are particularly effective. Table 3 [69] gives the equivalent thicknesses of lead and other materials for X rays excited at voltages from

Table 1

X RAYS GENERATED BY PEAK VOLTAGES	Minimum equivalent thickness of lead
Not exceeding kv.	mm.
75	1·0
100	1·5
125	2·0
150	2·5
175	3·0
200	4·0
250	6·0
300	9·0
350	12·0
400	15·0
(600)	(35·0)

150 kv. to 400 kv. peak and for radium γ rays. It is interesting to note that, as the exciting voltage of the radiation is increased, light materials become progressively more effective as compared with lead, and in the case of γ rays the relative lead equivalents are nearly proportional to the densities of the substances.

When handling sealed radium containers, care must be taken to guard against the effects of secondary β rays. In general it is not possible to provide absorbing screens, but ill effects can usually be avoided by expeditious working, and by handling the containers with long-handled forceps. In this connection a compromise must be made, since forceps with very long handles are apt to be somewhat clumsy, and any gain due to larger working distances may well be offset by the longer time involved in the manipulations.

Unsealed radioactive substances should be handled as little as possible, and thin rubber gloves should be worn when there is possibility of the substances coming in contact with the skin. Special care should be taken to

Table 2. *Lead protection required to secure tolerance dose at various distances from various radium sources. Tolerance dosage rate taken as* 10^{-5} *röntgen per second*

Quantity of radium (0·5 mm. Pt. screen)	Thickness of lead to give tolerance dose at the following distances from radium source							Tolerance distance with no lead
	Close up	25 cm.	50 cm.	1 metre	2 metres	5 metres	10 metres	
gm.	cm.	cm.	cm.	cm.	cm.	cm.	cm.	metres
0·05	9·5	5·5	2·5	0·5	—	—	—	1
0·2	11·5	8·5	5·5	2·5	0·5	—	—	2
0·5	13·0	10·5	7·5	4·5	2·0	—	—	3·5
1·0	14·5	12·0	9·0	6·0	3·0	—	—	4·5
2·0	15·5	13·5	10·5	7·5	4·5	1·0	—	6·5
5·0	17·0	15·5	12·5	9·5	6·5	2·5	0·5	10·5
10·0	18·0	17·0	14·0	11·0	8·0	4·0	1·5	14·5

Table 3. *X-ray lead equivalents of various materials*

(Density of lead 11·4 gm./c.c.)

Material	Mean density	Lead equivalent	Equivalent thickness of material				
			X rays excited at peak voltage of				Radium gamma rays
			150 kv.	200 kv.	300 kv.	400 kv.	
	gm./c.c.	mm.	mm.	mm.	mm.	mm.	mm.
Iron	7·9	1	11	12	12	11	2·5
		2	25	27	20	18	5
		3	37	40	28	23	7
		4	50	55	35	28	8·5
		6			48	38	12
		8			60	45	16
		10			75	55	19
		15				75	27
		20					35
		50					80
		100					150
Barium concrete or plaster—	3·2	1	10	14	14	13	5
2 parts coarse $BaSO_4$		2	21	30	27	24	10
2 parts fine $BaSO_4$		3	35	45	40	35	15
1 part cement		4	50	60	50	45	19
or		6			70	65	28
1 part coarse $BaSO_4$		8			90	80	36
1 part fine $BaSO_4$		10			120	100	45
1 part cement		15				140	65
		20					85
		50					200
		100					400
Barium concrete or plaster—	2·7	1	17	22	22	18	7
2 parts coarse $BaSO_4$		2	38	50	42	35	13
2 parts sand		3	65	75	60	50	18
1 part cement		4	90	100	75	60	24
or		6			105	85	34
1 part coarse $BaSO_4$		8			135	110	45
1 part sand		10			165	130	55
1 part cement		15				185	75
		20					95
		50					225
		100					450

Table 3 (*cont.*)

Material	Mean density	Lead equivalent	Equivalent thickness of material				
			X rays excited at peak voltage of				Radium gamma rays
			150 kv.	200 kv.	300 kv.	400 kv.	
	gm./c.c.	mm.	mm.	mm.	mm.	mm.	mm.
Concrete	2·2	1	85	80	80	50	8
2 parts ballast		2	160	150	95	75	16
2 parts sand		3	230	210	125	100	22
1 part cement		4	295	275	150	120	29
		6			210	150	40
		8			260	185	55
		10			300	220	65
		15				300	90
		20					120
		50					280
		100					550
Brick	1·9	1	110	100	85	80	11
(Daneshill red)		2	200	190	140	110	20
		3	280	270	170	140	28
		4	370	350	210	160	37
		6			280	210	50
		8			340	260	65
		10			400	300	80
		15				400	110
		20					150
Brick	1·6	1	130	130	100	90	15
(Yellow stock)		2	240	240	150	130	27
		3	340	340	200	160	39
		4		430	240	180	50
		6			320	240	70
		8			390	290	85
		10			460	340	105
		15				450	150
		20					200
Coke breeze	1·2	1	140	150	120	110	16
		2	250	270	190	160	30
		3	350	380	240	200	43
		4		490	290	230	55
		6			380	300	75
		8			460	350	95
		10			550	400	115
		15				510	165
		20					220

avoid inhaling particles of radioactive substances on account of the possibility of radium poisoning (see below).

Ventilation

Unless very carefully designed, bare conductors raised to high potentials are apt to give rise to electrical brush discharge, which causes chemical changes in the air giving rise to ozone and various oxides of nitrogen. This combination of "noxious fumes" is well known to cause sickness and headaches, and extended over a period, to general debility, so that X-ray departments must be provided with adequate artificial ventilation, preferably by the use of suitable exhaust fans. In the case of apparatus giving rise to corona, ten changes of air per hour is considered sufficient, while if the apparatus is corona-free six changes of air per hour are adequate for rooms of normal sizes (say 3000 cu. ft.). In general, modern apparatus of the "shock-proof" type is corona-free.

Electrical risks in X-ray departments

High-tension generators for X-ray purposes are usually supplied with power at commercial voltages and frequencies, while their output voltages range from 50 kv. to 400 kv. or even more in exceptional circumstances. Modern apparatus is usually of a shock-proof type in which the high-tension circuit is completely enclosed in earthed conductors. In some cases the high-tension transformer and X-ray tube are mounted together in a single earthed metal tank, while in other cases use is made of high-tension cables in which the conductor is surrounded by thick rubber insulation and a flexible earthed metal sheath.

§ 7. RADIUM POISONING

Although the need for the utmost care in the manipulation of radioactive substances is well known to physicists and others who have to deal with such preparations, the following brief reference to what is generally termed "radium poisoning" may not be without interest. This type of poisoning was first recognized in 1924, in workers employed in the painting of watch and instrument dials with radioactive luminous paint in the United States. The occurrences were traced to the practice of pointing the brush carrying the radioactive paint with the mouth, although evidence [70] was later obtained which suggested that the inhalation of radium dust and radon could also be a contributory factor. Other practices which have resulted in radium poisoning are the injection of radium solutions into the blood stream for therapeutic purposes and the drinking of liquid preparations containing radium. Colwell and Russ [71] and Loeb [72] have stressed the danger of the latter practice.

Perhaps the most serious characteristic of radium poisoning is that the symptoms (necrosis of the jaw, low blood pressure, regenerative anaemia, etc.) are slow in appearing and may not be discerned until some years after the radium has entered the body. Investigations have shown that although a large fraction of the radium taken by the mouth or by injection is eliminated within a few days, the rate of elimination rapidly decreases, finally becoming

negligible. A fraction of the radium remains permanently in the body, being distributed mainly in the skeleton. Attempts to effect the removal of such radium from the body have so far met with little success. In fatal cases quantities ranging from 2 to 180 microgrammes [73] have been detected in the entire body. A small fraction of a microgramme deposited in the lung was considered by Rajewsky [74] to have caused the fatal poisoning of a worker employed in a radium plant. Read and Mottram [75] have referred to instances which suggest that the regular continued inhalation of small concentrations of radon may cause ill health. These authors have deduced a theoretical atmospheric tolerance concentration of radon, the value obtained being about 4×10^{-9} curie per c.c. of air. Experiments on mice by the same authors indicated, however, that the tolerance concentration for these animals lay between 5×10^{-10} and 10^{-10} curie per c.c. of air, being nearer the lower than the higher value, and they deduce that, allowing a safety factor of 10, the tolerance atmospheric concentration of radon for the human body would seem to be about 10^{-11} curie per c.c. of air.

In view of the insidious nature of radium poisoning and the small quantities of material involved, it is essential that measures should be taken to safeguard the health of all persons whose duties expose them to naked radium preparations. Suggestions having this object in view were made following an investigation [70] of the working conditions in the radium dial-painting factories in the U.S.A. They included periodic tests of the radio-activity of the air of the work-room and of the workers themselves, and stressed the necessity for adequate ventilation and extreme cleanliness in the factory and for thorough personal cleanliness, particularly as regards the hands and finger-nails, of the workers. The adoption of these and other measures prevented the occurrence of further serious cases.

The mining of pitchblende involves danger to health due to the creation of dust arising from drilling and blasting and to emanation leaking from the ore-bed, and the same dangers exist in the drying and grinding of the ore. In the extraction processes, large quantities of emanation are evolved both during the chemical stages and during the fractional crystallization which follows. Leitch [76] has summarized the clinical information bearing on the above and has laid down measures for minimizing the dangers and for ensuring the early detection of the symptoms of radium poisoning among workers in the industry. The International Recommendations for X-ray and Radium Protection (1937) and the Recommendations of the British X-ray and Radium Protection Committee (1938) include clauses whose aim is to prevent the inhalation of radon during manipulation of radon plants and from unsuspected leakage from radium containers.

REFERENCES

High-voltage apparatus

(1) Brasch and Lange. *Naturwissenschaften*, **18**, 765 (1930); *Z. Phys.* **70**, 10 (1931).

(2) Joliot, Lazard and Savel. *C.R. Acad. Sci., Paris*, **201**, 826 (1935).

(3) LAURITSEN and BENNETT. *Phys. Rev.* **32**, 850 (1928).
(4) LAURITSEN and CASSEN. *Phys. Rev.* **36**, 988 (1930).
(5) CRANE, LAURITSEN and SOLTAN. *Phys. Rev.* **44**, 514 (1933).
(6) CRANE. *Phys. Rev.* **52**, 11 (1937).
(7) COCKCROFT and WALTON. *Proc. Roy. Soc.* A, **136**, 619 (1932).
(8) COCKCROFT and WALTON. *Proc. Roy. Soc.* A, **144**, 704 (1934).
(9) ALLIBONE. *Engineer, Lond.*, **163**, 452 (1937).
(10) BOUWERS. *Strahlentherapie*, **63**, 537 (1938).
(11) BOUWERS and KUNTKE. *Z. tech. Phys.* **18**, 209 (1937).
(12) VAN DE GRAAFF. *Phys. Rev.* **38**, 1919 (1931).
(13) VAN ATTA, NORTHRUP, VAN ATTA and VAN DE GRAAFF. *Phys. Rev.* **49**, 761 (1936).
(14) VAN ATTA and NORTHRUP. *Amer. J. Roentgenol.* **41**, 633 (1939).
(15) HERB, PARKINSON and KERST. *Phys. Rev.* **51**, 75 (1937).
(16) See GRIMMETT. *Nature, Lond.*, **141**, 311 (1938).
(17) TRUMP and VAN DE GRAAFF. *J. Appl. Phys.* **8**, 602 (1937).
(18) CHARLTON, WESTENDORP, DEMPSTER and HOTALING. *J. Appl. Phys.* **10**, 374 (1939).
(19) TRUMP and VAN DE GRAAFF. *Phys. Rev.* **55**, 1160 (1939).
(20) SLOAN. *Phys. Rev.* **47**, 62 (1935).
(21) COCKCROFT. *J. Sci. Instrum.* **16**, 37 (1939).
(22) LAWRENCE and LIVINGSTON. *Phys. Rev.* **37**, 1707 (1931).
(23) CHADWICK. *Nature, Lond.*, **142**, 630 (1938).
(24) THOMAS. *Phys. Rev.* **54**, 580 (1938).
(25) MANN. *Nature, Lond.*, **143**, 583 (1939).
(26) SLOAN and LAWRENCE. *Phys. Rev.* **38**, 2021 (1931).
(27) SLOAN and COATES. *Phys. Rev.* **46**, 539 (1934).

Radium

(28) PERRY. *Brit. J. Radiol.* **9**, 648 (1936).
(29) KAYE, PERRY and JONES. *Brit. J. Radiol.* **10**, 282 (1937).
(30) GORSKOV and IONOV. *C.R.* (*Doklady*) *Acad. Sci., U.R.S.S.*, **17**, 1–2, 15 (1937).
(31) HALBAN. *C.R. Acad. Sci., Paris*, **206**, 1170 (1938).
(32) *Rep. Progr. Phys.* **2**, 313 (1935) ; **3**, 366 (1936).
(33) POOLE. *Sci. Proc. R. Dublin Soc.* **21**, 595 (1938).
(34) POOLE. *Sci. Proc. R. Dublin Soc.* **21**, 608 (1938).
(35) EVANS and MUGELE. *Rev. Sci. Instrum.* **7**, 441 (1936).
(36) EVANS. *Amer. J. Roentgenol.* **37**, 368 (1937).
(37) CURTISS. *Bur. Stand. J. Res., Wash.*, **21**, 779 (1928) ; **23**, 137 (1939).

X-ray and γ-ray dosimetry

(38) *Rep. Progr. Phys.* **1**, 236 (1934) ; **2**, 309 (1935) ; **5**, 284 (1938).
(39) KAYE and BINKS. *Proc. Roy. Soc.* A, **161**, 564 (1937).
(40) MAYNEORD and ROBERTS. *Brit. J. Radiol.* **10**, 365 (1937).
(41) GRAY. *Brit. J. Radiol.* **10**, 600 (1937) ; **10**, 721 (1937).
(42) FAILLA. *Radiology*, **29**, 202 (1937).
(43) FAILLA and MARINELLI. *Amer. J. Roentgenol.* **38**, 312 (1937).
(44) FRIEDRICH, SCHULZE and HENSCHKE. *Strahlentherapie*, **60**, 38 (1937).
(45) LAURENCE. *Canad. J. Res.* **15**, 67 (1937).
(46) LAURITSEN. *Brit. J. Radiol.* **11**, 471 (1938).
(47) QUIMBY. *Radiology*, **32**, 583 (1939).
(48) HOLTHUSEN. *Radiološki Glasnik* (1938).
(49) NAIDU. *Amer. J. Roentgenol.* **41**, 992 (1939).

(50) Gray. *Proc. Roy. Soc.* A, **159**, 263 (1937).
(51) Zlotowski. *J. Phys. Radium*, **6**, 242 (1935).

Neutron dosimetry

(52) Chadwick. *Nature, Lond.*, **142**, 630 (1938).
(53) Lawrence, J. H. and Lawrence, E. O. *Proc. Nat. Acad. Sci., Wash.*, **22**, 124 (1936).
(54) Lawrence, E. O. *Radiology*, **29**, 313 (1937).
(55) Aebersold. *Phys. Rev.* **51**, 375 (1937).
(56) Zirkle and Lampe. *Amer. J. Roentgenol.* **39**, 613 (1938).
(57) Timoféeff-Ressovsky and Zimmer. *Naturwissenschaften*, **26**, 362 (1938).
(58) Zimmer. *Strahlentherapie*, **63**, 517 (1938).
(59) Zimmer, Born and Wolf. *Strahlentherapie*, **65**, 444 (1939).
(60) Gray and Read. *Nature, Lond.*, **144**, 439 (1939).
(61) Zimmer and Timoféeff-Ressovsky. *Strahlentherapie*, **63**, 528 (1938).
(62) Gray and Read. *Nature, Lond.*, **144**, 509 (1939).

X-ray and radium protection

(63) Mutscheller. *Amer. J. Roentgenol.* **13**, 65 (1925).
(64) Barclay and Cox. *Amer. J. Roentgenol.* **19**, 551 (1928).
(65) Russ and Scott. *Brit. J. Radiol.* **10**, 619 (1937); **12**, 39 (1939).
(66) *Recommendations of the British X-ray and Radium Protection Committee* (Fifth Revised Report, January, 1938).
(67) *International Recommendations for X-ray and Radium Protection* (September, 1937).
(68) Kaye, Bell and Binks. *Brit. J. Radiol.* **8**, 6 (1935).
(69) Kaye, Binks and Bell. *Brit. J. Radiol.* **11**, 676 (1938).

Radium poisoning

(70) Schwartz and others. *J. Industr. Hyg.* **15**, 362 & 433 (1933).
(71) Colwell and Russ. *Lancet*, **2**, 221 (1932).
(72) Loeb. *J. Amer. Water Works Ass.* **26**, 461 (1934).
(73) Evans. *Amer. J. of Public Health*, **23**, 1017 (1933).
(74) Rajewsky. *Radiology*, **32**, 57 (1939).
(75) Read and Mottram. *Brit. J. Radiol.* **12**, 54 (1939).
(76) Leitch. *Bull. Nat. Res. Coun., Ottawa*, no. 16 (1935).

THE CYCLOTRON AND SOME OF ITS APPLICATIONS

By W. B. MANN.
Imperial College, London

§ 1. INTRODUCTION

UNTIL now, no article in these *Reports* has been devoted to the subject of the Cyclotron. On the other hand there are, at the present time, few people who are not to a greater or less degree acquainted with the work which is going on in the cyclotron laboratories of the world. It is therefore difficult to decide, in the first place, where this article should be started and, secondly, in what detail it should be treated. It seems probable, however, that a brief survey of the whole subject, serving as a kind of skeleton on which to hang as complete a bibliography as possible, will best serve the purpose of these *Reports*.*

§ 2. MAGNETIC RESONANCE ACCELERATION

Initial and successful attempts had been made [1], [2], [3] to obtain heavy positive ions of high energy by successive accelerations through relatively low applied voltages.† This method, the method of linear resonance acceleration, was probably carried to its practical limit in the attainment [4] of a current of one-hundredth of a microampere of mercury ions having final energies equal to 2,850,000 electron-volts using an applied voltage of no more than 79,000 volts. For the purposes of nuclear physics, however, it is of more interest to be able to produce high-energy bombarding particles of low atomic number. To apply the method of linear resonance acceleration to, say, the production of high-energy protons would, however, necessitate the construction of an accelerating tube of such a length as to cause not a little trepidation before embarking upon the task. Thus, using frequencies of the order of 10^7 cycles per second it would require an accelerating tube some 25 metres long to produce protons of 3 million-electron-volts energy. While the construction of such an accelerating tube is not necessarily a practical impossibility, difficulties become apparent when one starts thinking not in terms of an upper limit of 3 Mev. but of, say, 10 or 20 Mev. protons. To circumvent this difficulty Lawrence [5] proposed the use of a magnetic field by means of which ions of small atomic mass could be caused to travel in semicircular paths within two

* Some of the material which has been used in preparing this *Report* was originally collected for a monograph on the Cyclotron and is here used again by kind permission of the publishers, Messrs. Methuen and Co., Ltd.

† This matter is also referred to in the *Report* by Kaye, Bell, Binks and Perry, p. 100. [EDITOR.]

hollow electrodes, or " dees " as they are now called, across which a high-frequency oscillating voltage is applied. The magnetic field is adjusted until the ions are in resonance with the high-frequency electric field between the dees—that is to say, until the ions move in semicircular paths of such a radius that each time they traverse the gap between the dees the direction of the electric field is such as still further to increase their energy. The fact that resonance can be maintained for ions of continuously increasing energy follows directly from Larmor's theorem, the angular velocity of an ion in a magnetic field being independent of its linear velocity, that is, independent of its energy. The time taken for an ion to complete the " semicircular " * path within the dees between two successive accelerations is therefore the same, whatever may be the energy of the ion and the radius of curvature of its path, so long as its linear velocity remains small compared with the velocity of light. The ions thus gain energy at each traversal of the gap between the dees and circulate in the relatively field-free region within the dees with ever-increasing radius of curvature of their semicircular paths until, finally, they emerge at the edge of one of the dees with energies corresponding to accelerations through several million volts. The potentialities of the magnetic resonance accelerator or " cyclotron ", as it is now more usually called, have been more than adequately demonstrated [6]–[13].

§ 3. ENERGY OF EMERGENT ION BEAM

The energy of an ion, of mass m and charge e, which is circulating within the dees of the cyclotron along a path having a radius of curvature ρ is such as to correspond to a total acceleration through an equivalent voltage V given by

$$V = 16{\cdot}7 \times 10^{-20} H^2 \rho^2 e/m, \qquad \ldots.(1)$$

where H is the intensity of the magnetic field, this relation giving V in volts when the substituted value of e is in electrostatic units.

The value of H in equation (1) involves in turn the frequency (n) of oscillation of the high-frequency voltage applied to the dees. Thus suppose that ω is the angular velocity of the ion which by Larmor's theorem is, for a given e/m, a function only of the magnetic field H. The time taken for an ion to complete a " semicircle " within the dees is therefore equal to π/ω, while the time required to complete two such " semicircles " consecutively should be equal to $1/n$ if the ion is always to cross the gap between the dees in phase with the electric field. Thus for resonance we must have :

$$2\pi/\omega = 1/n. \qquad \ldots.(2)$$

* The doubt associated with the word " semicircular ", both here and later in the text, by the use of quotation marks is on account of the fact that only in the ideal case where the acceleration occurs over an infinitesimally small distance will consecutive paths within the dees be complete semicircles. In actual fact the electric field penetrates within the dees and, over a finite length of path where the ion is being accelerated, the radius of curvature increases continuously.

But the equation of motion of an ion, of mass m and charge e, moving along a path of radius of curvature ρ with a velocity v in a magnetic field of intensity H is given by the relation :

$$Hev = mv^2/\rho, \qquad \ldots.(3)$$

which in conjunction with equation (2) gives for the condition of resonance :

$$n = (H/2\pi).(e/m), \qquad \ldots.(4)$$

since $\omega = v/\rho$.

This gives the value of the magnetic field (H) required to give resonance for ions of given e/m for any given frequency (n) of the high-frequency voltage applied to the dees. In practice, these frequencies are radio-frequencies of the order of 10^7 cycles per second. The higher the frequency the higher is the value of the magnetic field required to give resonance, and hence an upper limit to the frequency is imposed by the limit set to the magnetic field on account of saturation of the iron of the electromagnet. As will be seen from equation (1), the higher the frequency and the higher the magnetic field, the greater will be the energy of the ions, of given e/m, circulating at any given radius from the centre of the dee system.

Usually the dee circuit is tuned to a wave-length between 20 and 30 metres, that is to frequencies between $1{\cdot}5 \times 10^7$ and $1{\cdot}0 \times 10^7$ cycles per second, respectively. Substituting the latter value of frequency into equation (4), we find that the values of the magnetic field required to give resonance are 6,500 and 13,000 gauss in the cases of protons and deuterons respectively. At a frequency of $1{\cdot}5 \times 10^7$ cycles per second, the resonance values of the magnetic field will have been increased by 50 per cent and will have reached, in the case of deuterons, the maximum allowed by the saturation of the iron. Nor in the case of alpha-particles, having approximately the same value of e/m as deuterons, will it be possible to use radio-frequencies much greater than $1{\cdot}0 \times 10^7$ cycles per second.

By inspection of equation (4) it is seen that, an upper limit having been set to the equivalent voltage (V) on account of the magnetic field (H), it is still possible to increase the equivalent voltage by increasing the dimensions of the cyclotron and so allowing the ions to circulate within the dees at much greater radii of curvature (ρ). There is, therefore, a tendency to-day towards the construction of cyclotrons of large dimensions, and the completion and successful operation has recently been reported [14] of the largest cyclotron which has yet been built, the new 60-inch cyclotron at Berkeley. With this cyclotron, deuterons having final energies equal to 19 million electron-volts have been obtained. It would perhaps be of interest to consider how the final energy of the emergent ions is dependent upon the frequency of the applied voltage and the intensity of the magnetic field in such a cyclotron. To do this it is necessary to substitute, for the radius of curvature (ρ) in equation (1), the radius of curvature of the ions as they emerge from the dees, that is the distance between the exit slit and the centre of the dee system. No such dimension has yet been published for the new 60-inch cyclotron,

but we may take, say, 60 cm. as an approximate figure. Substituting therefore, $\rho=60$ cm. in equation (1) and the appropriate values of e and m, we obtain the curves shown in figure 1, in which the equivalent voltage (V) of the emergent ion beam is plotted as a function of the magnetic field (H). Also shown are the values of the frequency (n) of the radio-frequency voltage applied to the dees, which values are related to the magnetic field (H) by equation (4). For the higher energies the mass (m) in equation (1) has been replaced by the relativistic mass $m/\sqrt{\{1-(v/c)^2\}}$, where m is now the rest-mass of the ions. This correction is, however, still small within the range of the curves shown in figure 1, being equal to about one per cent for a 15 million-electron-volt

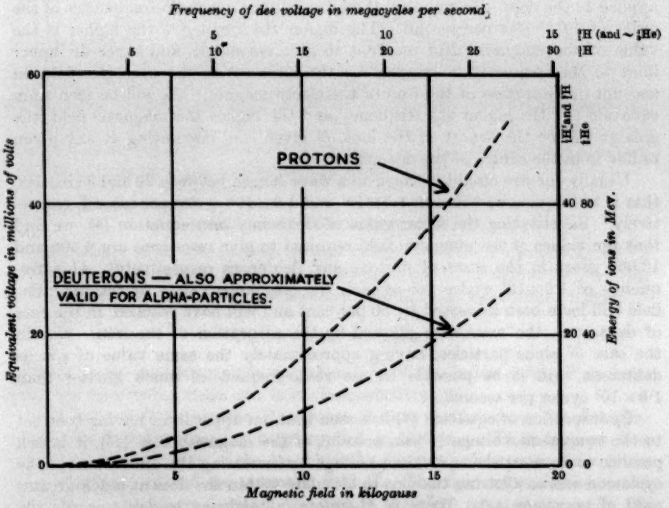


Figure 1. Equivalent voltage of emergent ions at various fields.

deuteron and three per cent for a deuteron having an energy equal to 27 million electron-volts.

The deuteron curve of figure 1 is also approximately applicable to the case of doubly ionized helium atoms, the value of e/m which is to be substituted in equation (1) in the case of alpha-particles being approximately one-half per cent larger than the value of e/m for deuterons. The energy of the emergent alpha-particle beam in electron-volts will, however, be twice the numerical value of the final equivalent voltage (V).

§ 4. EFFECT OF RELATIVITY

Several papers have now been published [15], [16], [17] in which the effect of relativity, in imposing an upper limit to the energy of the ions obtainable from the cyclotron, has been discussed. When the linear velocities of the ions become comparable with the velocity of light there will be a relativistic increase in mass which will cause an increase in the radius of curvature of the "semi-circular" path within the dees. The distance travelled by these high-energy ions between consecutive accelerations in the gap between the dees will consequently be increased and the ions will tend to lag behind the accelerating electric field. A magnetic field which increased at large radii would cause a decrease in the radius of curvature of the path travelled by the high-energy ions and would therefore compensate for the relativity effect. Such a magnetic field would, however, cause a defocussing of the ion beam and consequent loss in intensity. Thomas [17] proposed, however, to introduce a periodic variation of magnetic field along the path of the ion, which would introduce a further focussing effect to compensate for the defocussing caused by the increase in the magnetic field at large radii. It seems, however, that for the present a rather simpler expedient will suffice, namely to increase the voltage applied to the dees as in the case of the new 60-inch cyclotron at Berkeley. For an ion to be lost from the beam it must arrive at the accelerating gap between the dees with such a time lag behind the electric field that it finds not an accelerating, but a retarding, field in the gap. For this to occur an ion would have to be out of phase and behind the electric field by an amount greater than $\pi/2$. For, say, a one or two per cent relativistic increase in mass there will be a corresponding lag in phase of from two to four degrees at each traversal of the accelerating gap. Only after some twenty or thirty such traversals will the total phase lag approach $\pi/2$, bearing in mind, however, that the higher the energy of the ion becomes, the greater is the percentage increase in mass, due to relativity, over the rest-mass of the ion. By increasing the dee voltage, the ions acquire their final energies with a smaller number of revolutions within the dees and with fewer traversals of the accelerating gap between the dees. The small phase lag corresponding to each acceleration is thus not allowed to build up to such a value that the ions would become more than $\pi/2$ out of phase with the electric field and consequently lost from the beam.

§ 5. MAGNETIC AND ELECTROSTATIC FOCUSSING

Some very careful investigations have been made by Rose [18] and Wilson [18] into the operation of magnetic and electrostatic focussing in the cyclotron. One of the most important results of these investigations is the advocacy by Wilson of a magnetic field which decreases continuously from the central region between the poles to the edge of the pole-faces. Such a field having a small radial decrease greatly enhances the magnetic focussing effect. In addition it means that most of the ions finally emerging from the edge of the dee must have started life lagging in phase behind the dee voltage. But it has been pointed out by Rose that the electrostatic focussing effect is a maximum for such ions—ions which cross the accelerating gap at an instant when the

potential difference between the dees is decreasing. The result of introducing a slight radial decrease into the magnetic field is thus to increase the effects both of magnetic and electrostatic focussing. Such a radial gradient can be introduced either by bevelling the pole-faces of the electromagnet or, more simply, by inserting thin circular plates of iron (circular "shims", as they are now usually called) of gradually increasing radius into the space between the pole-faces and the bottom and lid of the vacuum chamber. The adoption of this latter expedient at Berkeley greatly increased the intensity of the beam of high energy ions from the 37-inch cyclotron.

§ 6. THE RADIO-FREQUENCY DEE VOLTAGE SUPPLY

In the earlier cyclotrons two power valves operating in a self-excited, tuned-grid, tuned-plate, push-pull circuit were utilized to supply radio-frequency power to the dee circuit. The power valves used in this case were usually of a type which has been described by Sloan, Jenkins and Thornton [19]. The construction of such valves is relatively straightforward. With the advent, however, of devices for stabilizing the magnetic field [20], [12] it became advantageous to maintain more nearly constant the frequency of the radio-frequency supply to the dees. This improvement in frequency stability can be obtained by the use of circuits of so-called high Q. A quartz crystal oscillator is equivalent to a circuit of high Q and the first development in the direction of increased stability of frequency lay in the use of a master-oscillator, the frequency of which was controlled by means of a quartz crystal. The output of the master-oscillator is passed through a number of stages of voltage amplification and is then fed into a power amplifier, the output from which is fed into the dee circuit of the cyclotron.

At high frequencies another circuit which may conveniently be used to give frequency stability, and which has a very high Q, is the resonant quarter-wave line. At frequencies of the order of 10^7 cycles per second, a resonant quarter-wave line has a Q which is higher than that of any but the most elaborately mounted quartz crystals. The properties of these line circuits have been discussed in detail [21]–[26].* Their use in the form of resonant quarter-wave lines allows of a return to the simple self-excited, tuned-grid, tuned-plate, push-pull oscillator, thereby eliminating much of the expense necessitated by the use of a master-oscillator with many stages of amplification. Such a return is to be noted in many cyclotron laboratories, a resonant quarter-wave line being incorporated in the grid circuit, and possibly also in the plate circuit, of two power valves arranged in push-pull. Resonant lines, both open and concentric, have also been incorporated in the dee circuits at Berkeley and elsewhere [27] The use of non-resonant transmission lines [10] for feeding power from the oscillators into the dee circuit is now also a widely adopted practice. The use of a transmission line enables the oscillators to be placed at a convenient distance from the cyclotron itself and eliminates the crowding inherent in the older cyclotrons, where the oscillators were usually coupled inductively to the dee circuit.

* They are also discussed in the *Report* by Hartshorn, p. 378. [EDITOR.]

§7. THE ION SOURCE

An important contributing factor in the attainment of large currents of high-energy ions from the cyclotron is the type of ion source employed. In the earlier cyclotrons this usually consisted of a heated tungsten filament situated at the centre of the vacuum chamber either above or below the dees. Electrons emitted from this filament were then accelerated by putting the filament assembly at a negative potential of a few hundred volts with respect to the vacuum chamber. These electrons then travelled vertically along the lines of magnetic force, in reality performing very close spirals around the lines of force, and, by collision with the neutral molecules, provided a vertical column of positive ions at the centre of the chamber. From this column positive ions were drawn into the dees.

In the early days the heating of the tungsten filament was effected by means of direct current. Using direct current, however, in such an intense magnetic field resulted in the filaments being subjected to quite considerable strains and the lives of these filaments were consequently of short duration. On this account it is now customary to use high-frequency heating when a tungsten-filament type of ion source is employed [28].

The current of high-energy ions emerging from the cyclotron is directly dependent upon the number of ions produced at the centre of the dees, that is upon the efficiency of the ion source. The efficiency of the ion source is, however, dependent in turn upon three factors, namely, the number and energy of the ionizing particles and the pressure of the gas in the vacuum chamber. The number of ionizing electrons should be as large as possible and, in the case of a tungsten filament, the emitting surface should therefore be large, and is usually in the form of a flat spiral. The energy of the ionizing electrons is found to be an optimum when they have been accelerated through a potential of about 200 volts [9], [29], [30]. The gas pressure in the vacuum chamber is usually determined by considerations other than that of obtaining optimum conditions for ion production. An upper limit is set to the pressure by the fact that, with such large potentials on the dees, electrical breakdown will occur should the pressure become too high. When electrical discharges occur, then the oscillators become overloaded and the current of high-energy ions rapidly decreases.

The recent introduction [31], [32] of capillary ion sources for use with the cyclotron represents a considerable advance and obviates a number of difficulties. The gas pressure can be adjusted for optimum conditions of ion production while at the same time the differential pressure between the capillary source and the vacuum chamber can be considerable. Originally the whole vacuum chamber had to be filled with gas ; with the use of a capillary ion source, however, the gas can be introduced into the source itself while powerful pumps keep the pressure in the vacuum chamber as low as possible, thereby preventing electrical discharges while, at the same time, the probability of ions being scattered from the beam is also diminished. In a recent letter, McMillan and Salisbury [33] report that, by the introduction of a capillary arc source, the beam of 16 Mev. deuterons from the 60-inch cyclotron at

Berkeley was increased from 14 to 90 microamperes. Another advantage of this type of source over the more distended vertical column of ions, which has been pointed out by Livingston, Holloway and Baker [31], is that it is possible to use dees of much smaller vertical aperture on account of the high-energy ions being more concentrated in a horizontal sheet. This allows of a smaller gap between the magnet poles and a consequent saving in power.

§ 8. SOME APPLICATIONS

A prolific source of high-energy protons, deuterons and alpha-particles and of neutrons,* the cyclotron has been successfully applied to a very wide range of scientific investigation. In particular has it been used in the preparation and identification of radioactive isotopes of the stable elements [34], the large yields from the cyclotron greatly facilitating the work of chemical identification.† As a corollary to this is the preparation of artificial radioactive substances for indicator experiments, a task for which the cyclotron is again ideally suited. Particularly has the Berkeley cyclotron been used for the preparation of radioactive phosphorus ($^{32}_{15}$P) [35] as an indicator in experiments on exchanges of phosphorus in plants [36], [37] and on phosphorus metabolism [38]–[42].

Of particular interest, in view of the desirability of obtaining as large yields as possible of radioactive material, was the investigation by Wilson [43] of the currents of high-energy ions circulating within the dees. He found that only about 10 per cent of the available high-energy ions were being withdrawn from the dees and that ion currents of the order of a milliampere were circulating within the dees themselves. Subsequently Wilson and Kamen [44] have utilized these large currents for the preparation of radioactive iron ($^{59}_{26}$Fe) [45] and phosphorus, the former for use in studying iron metabolism in relation to anaemia [46], by inserting an iron phosphide target between the dees. Previously, radioactive phosphorus had been prepared by bombardment of red phosphorus in a specially constructed target chamber [47] which allowed of the phosphorus being maintained in an atmosphere of either helium or hydrogen, at a pressure of a few centimetres of mercury, during bombardment.

In addition to the experiments with phosphorus and iron, other radioactive elements have been extensively used in indicator experiments both in biology [38], [48] and chemistry [49]. The biological effects of neutrons, using the cyclotron as the source, have also been investigated [38], [50].

In physics, in addition to its very extensive use in the investigation of nuclear reactions, the cyclotron has been used by Alvarez [51] to provide a source of "monochromatic" neutrons. In his experiments Alvarez modulated the power input to the dee circuit by applying 60-cycle alternating voltage, after half-wave rectification, to the plates of the power amplifier. Deuterons were incident upon the beryllium target in bursts and corresponding bursts of neutrons were generated. A boron-trifluoride ionization chamber set at

* Lawrence and Cooksey [9] estimate the yield of neutrons ($^{9}_{4}$Be+$^{2}_{1}$H) from the 37-inch cyclotron at Berkeley to be more than 10^5 times as great as that from one curie of radon mixed with beryllium.

† Compare the *Report* by Walke, p. 16. [Editor.]

a distance (d) from the neutron source was arranged to be sensitive at a short time (t) after the arrival of the burst of deuterons at the beryllium target. This was done by adjusting the phase of the 60-cycle voltage, applied to the power amplifier plates, relative to the phase of the same frequency modulation which was applied to the linear amplifier into which the output from the boron-trifluoride ionization chamber was fed. The velocity of the neutrons recorded in the ionization chamber was then given by the ratio of the distance (d) to the time lag (t). The resolution in the neutron velocity spectrum obtained by Alvarez was not, however, so sharp as that obtained by the application of the same idea using deuterons accelerated in a discharge tube [52]. The experiments with the discharge tube were identical in principle, the essential difference in method being that shorter and sharper pulses of voltage were applied in the case of the discharge tube, thus giving increased sharpness of resolution in the neutron velocity spectrum.

An interesting experiment has recently been performed, by Alvarez and Cornog [53], in which the 60-inch cyclotron at Berkeley was used as a mass spectrograph in order to determine whether the isotope of helium of atomic mass three ($^3_2$He) might be a stable constituent of ordinary helium. A beam of high-energy ions, of 54-cm. range in air, was detected at a setting of the magnetic field which corresponded to $^3_2$He ions of the observed range, the setting of the magnetic field being mid-way between the two values corresponding to protons of 8 Mev. and alpha-particles of 32 Mev. The beam of $^3_2$He ions so obtained was, however, too unsteady for bombardment purposes. Subsequently, however, Alvarez and Cornog [54], by "shimming" adjustment to the magnetic field, have succeeded in obtaining a steady beam of "light alpha-particles" with which silicon was bombarded and an activity, ascribed to $^{30}_{15}$P, observed. At the same time these authors report experimental evidence of the existence of a radioactive hydrogen isotope of atomic mass three ($^3_1$H).

As a very powerful neutron source, the cyclotron has been used in many of the recent investigations into the fission of the uranium nucleus, but to none of these investigations will reference be made here since they more properly belong to the *Report* on induced radioactivity.*

For general purposes of bombardment, either for the preparation of radioactive sources or for the production of neutrons, it is no disadvantage if the high-energy ions are still moving in a curve of relatively small radius of curvature. For some purposes, however, as, for instance, for scattering experiments, it may be desirable that the ions should be travelling approximately linearly. This may be achieved by means of an arrangement which, in the *Physical Review* [10], has been referred to as a "beam exit appurtenance" but which, in cyclotron laboratories, is usually more familiarly called a "snout". This snout consists of two or three long tubes in series, evacuated and attached at one end to the target chamber of the cyclotron. Through these tubes the beam of high-energy ions passes with ever-increasing radius of curvature into regions of correspondingly decreasing magnetic intensity. Finally the beam emerges from the end of the snout at a point so far removed from the

* This volume, p. 16.

centre of the electromagnet that the magnetic field is negligibly small and the path of the ions sensibly linear. The longer the snout the more nearly will the ion paths become linear, but the smaller, also, will become the intensity of the beam. This intensity may, to some extent, be restored by the use of an auxiliary focussing magnet.

In accordance with the purpose set forth in its introduction, this *Report* has been kept as brief as possible, but a full bibliography has been given which, it is hoped, may be commensurate with the importance of the cyclotron and of its contributions to physics and other fields of scientific investigation. In many instances a group of references has been given which refers rather to a subject than to particular investigations in that subject. The references which now follow have therefore been expanded to include the titles of the individual papers, so that a general reference may at the same time be more specific to those who may desire it.

REFERENCES

(1) Wideröe, R. "On a new principle for the attainment of high voltages." *Arch. Elektrotech.* **21**, 387 (1929).

(2) Lawrence, E. O. and Sloan, D. H. "The production of high-speed canal rays without the use of high voltages." *Proc. Nat. Acad. Sci., Wash.*, **17**, 64 (1931).

(3) Sloan, D. H. and Lawrence, E. O. "The production of heavy high-speed ions without the use of high voltages." *Phys. Rev.* **38**, 2021 (1931).

(4) Sloan, D. H. and Coates, W. M. "Recent advances in the production of heavy high-speed ions without the use of high voltages." *Phys. Rev.* **46**, 539 (1934).

(5) Lawrence, E. O. and Edlefsen, N. E. "On the production of high-speed protons." *Science*, **72**, 376 (1930).

(6) Lawrence, E. O. and Livingston, M. S. "The production of high-speed light ions without the use of high voltages." *Phys. Rev.* **40**, 19 (1932).

(7) Lawrence, E. O. and Livingston, M. S. "The multiple acceleration of ions to very high speeds." *Phys. Rev.* **45**, 608 (1934).

(8) Livingston, M. S. "The magnetic resonance accelerator." *Rev. Sci. Instrum.* **7**, 55 (1936).

(9) Lawrence, E. O. and Cooksey, D. "On the apparatus for the multiple acceleration of light ions to high speeds." *Phys. Rev.* **50**, 1131 (1936).

(10) Kruger, P. G. and Green, G. K. "The construction and operation of a cyclotron to produce one million volt deuterons." *Phys. Rev.* **51**, 699 (1937).

(11) Kurie, F. N. D. "The cyclotron: a new research tool for physics and biology." *Gen. Elect. Rev.* **40**, 264 (1937).

(12) Henderson, M. C. and White, M. G. "The design and operation of a large cyclotron." *Rev. Sci. Instrum.* **9**, 19 (1938).

(13) Kurie, F. N. D. "Present-day design and technique of the cyclotron." *J. Appl. Phys.* **9**, 691 (1938).

(14) Lawrence, E. O., Alvarez, L. W., Brobeck, W. M., Cooksey, D., Corson, D. R., McMillan, E. M., Salisbury, W. W. and Thornton, R. L. "Initial performance of the 60-inch cyclotron of the William H. Crocker Radiation Laboratory, University of California." *Phys. Rev.* **56**, 124 (1939).

(15) Bethe, H. A. and Rose, M. E. "The maximum energy obtainable from the cyclotron." *Phys. Rev.* **52**, 1254 (1937).

(16) Rose, M. E. "Focusing and maximum energy of ions in the cyclotron." *Phys. Rev.* **53**, 392 (1938).

(17) THOMAS, L. H. "The paths of ions in the cyclotron." *Phys. Rev.* **54** (1938). I. Orbits in the magnetic field, p. 580. II. Paths in the combined electric and magnetic fields, p. 588.

(18) WILSON, R. R. "Magnetic and electrostatic focusing in the cyclotron." *Phys. Rev.* **53**, 408 (1938).

(19) SLOAN, D. H., THORNTON, R. L. and JENKINS, F. A. "A demountable power-oscillator tube." *Rev. Sci. Instrum.* **6**, 75 (1935).

(20) WYNN-WILLIAMS, C. E. "An automatic magnetic field stabilizer of high sensitivity." *Proc. Roy. Soc.* A, **145**, 250 (1934).

(21) EVERITT, W. L. *Communication Engineering* (1937).

(22) CONKLIN, J. W., FINCH, J. L. and HANSELL, C. W. "Methods of frequency control employing long lines." *Proc. Inst. Radio Engrs, N.Y.*, **19**, 1918 (1931).

(23) STERBA, E. J. and FELDMAN, C. B. "Transmission lines for short-wave radio systems." *Bell. Syst. Tech. Jour.* **11**, 411 (1932).

(24) TERMAN, F. E. "Resonant lines in radio circuits." *Elect. Engng, N.Y.*, **53**, 1046 (1934).

(25) REUKEMA, L. E. "Transmission lines at very high radio frequencies." *Elect. Engng, N.Y.*, **56**, 1002 (1937).

(26) HANSELL, C. W. and CARTER, P. S. "Frequency control by low power-factor line circuits." *Proc. Inst. Radio Engrs, N.Y.*, **24**, 597 (1936).

(27) DUNNING, J. R. and ANDERSON, H. L. "High-frequency systems for the cyclotron." *Phys. Rev.* **53**, 334 (1938). See also : *Note added in proof*, p. 136 below.

(28) DUNNING, J. R. and ANDERSON, H. L. "High-frequency filament supply for ion sources." *Rev. Sci. Instrum.* **8**, 158 (1937).

(29) BLEAKNEY, W. "The ionisation of hydrogen by single electron impact." *Phys. Rev.* **35**, 1180 (1930).

(30) TATE, J. T. and SMITH, P. T. "The efficiencies of ionisation and ionisation potentials of various gases under electron impact." *Phys. Rev.* **39**, 270 (1932).

(31) LIVINGSTON, M. S., HOLLOWAY, M. G. and BAKER, C. P. "A capillary ion source for the cyclotron." *Rev. Sci. Instrum.* **10**, 63 (1939).

(32) ALLEN, A. J. "Improvement and development of the cyclotron." *J. Franklin Inst.* **228**, 109 (1939).

(33) MCMILLAN, E. M. and SALISBURY, W. W. "A modified arc source for the cyclotron." *Phys. Rev.* **56**, 836 (1939).

(34) LIVINGSTON, M. S. and BETHE, H. A. "Nuclear physics, experimental." *Rev. Mod. Phys.* **9**, 245 (1937).

(35) NEWSON, H. W. "The radioactivity induced in silicon and phosphorus by deuteron bombardment." *Phys. Rev.* **51**, 624 (1937).

(36) HEVESY, G., LINDERSTRØM-LANG, K. and OLSEN, C. "Atomic dynamics of plant growth." *Nature, Lond.*, **137**, 66 (1936); and "Exchange of phosphorus atoms in plants and seeds." *Nature, Lond.*, **139**, 149 (1937).

(37) HEVESY, G., LINDERSTRØM-LANG, K. and NIELSEN, N. "Phosphorus exchange in yeast." *Nature, Lond.*, **140**, 725 (1937).

(38) LAWRENCE, J. H. "Artificial radioactivity and neutron rays in biology and medicine." *Handbook of Physical Therapy*, 1 (1938). Further references are to be found in the bibliography of this article.

(39) CHIEVITZ, O., and HEVESY, G. "Radioactive indicators in the study of phosphorus metabolism in rats." *Nature, Lond.*, **136**, 754 (1935); and "Studies on the metabolism of phosphorus in animals." *Det Kgl. Danske Vid. Selskab. Biol. Medd.* **13**, paper 9 (1937).

(40) COOK, S. F., SCOTT, K. G. and ABELSON, P. "The deposition of radio phosphorus in tissues of growing chicks." *Proc. Nat. Acad. Sci., Wash.*, **23**, 529 (1937).

(41) "Radioactive phosphorus as an indicator of phospholipid metabolism." Papers I to VII in *J. Biol. Chem.* (I) **122**, 169 (1937), by I. PERLMAN, S. RUBEN and I. L. CHAIKOFF. (II) **123**, 587 (1938), by B. A. FRIES, S. RUBEN, I. PERLMAN and I. L. CHAIKOFF. (III) **124**, 795 (1938), by C. ENTENMAN, S. RUBEN, F. W. LORENZ and I. L. CHAIKOFF. (IV) **126**, 493 (1938), by G. W. CHANGUS, I. L. CHAIKOFF and S. RUBEN. (V) **127**, 211 (1938), by I. PERLMAN and I. L. CHAIKOFF. (VI) **128**, 631 (1939), by H. B. JONES, I. L. CHAIKOFF and J. H. LAWRENCE. (VII) **128**, 735 (1939), by I. PERLMAN and I. L. CHAIKOFF.

(42) COHN, W. E. and GREENBERG, D. M. "Studies in mineral metabolism with the aid of artificial radioactive isotopes." *J. Biol. Chem.* (I) **123**, 185 (1938), "Absorption, distribution and excretion of phosphorus." (II) **130**, 625 (1939), "The influence of vitamin D on the phosphorus metabolism of rachitic rats."

(43) WILSON, R. R. "Magnitude of accelerated current in the cyclotron." *Phys. Rev.* **54**, 240 (1938).

(44) WILSON, R. R. and KAMEN, M. "Internal targets in the cyclotron." *Phys. Rev.* **54**, 1031 (1938).

(45) LIVINGOOD, J. J. and SEABORG, G. T. "Radioactive isotopes of iron." *Phys. Rev.* **54**, 51 (1938).

(46) HAHN, P. F., BALE, W. F., LAWRENCE, E. O. and WHIPPLE, G. H. "Radioactive iron and its metabolism in anaemia: its absorption, transportation and utilization." *J. Exp. Med.* **69**, 739 (1939).

(47) KURIE, F. N. D. "High intensity bombardment with fast particles." *Rev. Sci. Instrum.* **10**, 199 (1939).

(48) HAMILTON, J. G. "Rates of absorption of radioactive isotopes of sodium, potassium, chlorine, bromine and iodine in normal human subjects." *Amer. J. Physiol.* **124**, 667 (1938). Further references have been given by J. H. LAWRENCE [38].

(49) OLSON, A. R., HALFORD, R. S. and HORNEL, J. C. "The rearrangement of N-chloroacetanilide." *J. Amer. Chem. Soc.* **59**, 1613 (1937).

(50) LAWRENCE, E. O. "The biological action of neutron rays." *Radiology*, **29**, 313 (1937). Further references have been given by J. H. LAWRENCE [38].

(51) ALVAREZ, L. W. "The production of collimated beams of monochromatic neutrons in the temperature range 300°–10° K." *Phys. Rev.* **54**, 609 (1938).

(52) FERTEL, G. E. F., MOON, P. B., THOMSON, G. P. and WYNN-WILLIAMS, C. E. "Velocity distribution of thermal neutrons." *Nature, Lond.*, **142**, 829 (1938).

(53) ALVAREZ, L. W. and CORNOG, R. "He^3 in helium." *Phys. Rev.* **56**, 379 (1939).

(54) ALVAREZ, L. W. and CORNOG, R. "Helium and hydrogen of mass 3." *Phys. Rev.* **56**, 613 (1939).

Note added in proof

Since going to press, two papers have been published describing the cyclotrons at the Biochemical Research Foundation and at Purdue University. In these papers comprehensive constructional details are given of the cyclotrons proper and also of the high-frequency systems and arc ion-sources. Performance data are also included. The references to these papers are:

(55) ALLEN, A. J., SAMPSON, M. B. and FRANKLIN, R. G. "The cyclotron of the Biochemical Research Foundation." *J. Franklin Inst.* **228**, 543 (1939).

(56) HENDERSON, W. J., KING, L. D. P., RISSER, J. R., YEARIAN, H. J. and HOWE, J. D. "The Purdue cyclotron." *J. Franklin Inst.* **228**, 563 (1939).

SPECTROSCOPY

Edited by W. JEVONS, D.Sc., PH.D.

I. THE SPECTRA OF SUPERNOVAE

BY A. HUNTER, PH.D., F.R.A.S.
Royal Observatory, Greenwich

AMONG the more striking recent discoveries in astrophysics has been the recognition of supernovae—a class of objects whose rarity of occurrence is as striking as the stupendous nature of the outbursts which they cause when they do occur. Ordinary novae are common enough: they appear at the rate of twenty or thirty per annum both in our own galaxy and in others (such as the Andromeda nebula) which are near enough for such objects to be recognized. Supernovae, however, occur only at the rate of one or two per galaxy per millennium; none has been observed in our own galaxy in modern times, and when, in 1936, a systematic search for supernovae in external galaxies was started with the 18-inch Schmidt camera on Mt. Palomar, the first nine months' work yielded 300 photographs showing images of nearly 10,000 nebulae, in only one of which appeared a supernova. What these objects lose in frequency of occurrence, however, they make up for in luminosity. Ordinary novae themselves reach, on the average, an absolute magnitude of -6, some 10 times brighter than the brightest supergiant star; supernovae, however, reach -15 at maximum, 4000 times brighter even than this, and 80×10^6 times brighter than our sun. Only twenty supernovae are yet known, three-quarters of them by chance observation and the rest as the result of the systematic search, mentioned above, for bright outbursts in spiral nebulae.

In discussing the origin of these objects, their relation to ordinary novae, and their significance as regards stellar evolution in general, the astrophysicist turns first of all to spectroscopy for assistance. Observations of their light variation, their frequency of occurrence and their positions in the parent galaxies can of course give valuable information, but such information will be regarded more as auxiliary to that afforded by the spectra than as significant in itself. Detailed descriptions of spectra have been published for only seven supernovae, but of these, two (in IC 4182 and NGC 1003) have been so well observed [1] that the broad features of the spectra can now be described.

First of all it must be realized that although the brightest supernovae radiate at maximum as much light as the total integrated light of the spiral nebulae in which they appear, they are also at such great distances that they are apparently quite faint. Only one of the twenty known supernovae (S Andromedae, the great supernova of 1885) has reached naked-eye brightness, and the brightest of those observed since wide-aperture stellar spectrographs became available (SN IC 4182) reached only 8·5 visual magnitude at

maximum. Accordingly no great dispersion or resolving power is available. Even so, nothing remotely resembling a laboratory low-dispersion spectrum is actually observed in any supernova. The spectra—they are all identical in general character and similar in most details—consist, at all stages of the outburst, of broad, almost featureless emission bands * containing numerous indistinct details, many perceptible only on microphotometer tracings. Wave-lengths of these details and of the band maxima themselves have, with few exceptions, to be measured from a set of tracings showing the comparison spectrum together with that of the supernova. This procedure—in itself a source of error—must moreover be carried out on tracings from original spectrograms which are often on film base rather than on glass. The resulting errors of measurement may range from 3A. in the violet to 15A. in the red. The difficulty of interpreting the results will best be appreciated when it is remarked that even then the accuracy of the wave-lengths is limited mainly by the intrinsic diffuseness of the features of the spectrum rather than by technical difficulties of measurement.

In such circumstances detailed identifications cannot be expected. In fact, with two exceptions, no identifications of the wide emission bands have yet met with general acceptance, whilst certain assertions [2] that absorption lines due to H, He, N and O are present are open to question. In particular, the Pickering series of He II will account for possible faint lines near the positions of the Balmer lines. It is at any rate certain that there is no strong hydrogen spectrum such as is so conspicuous in ordinary novae. Indeed, in spite of the fact that a supernova appears to differ from a nova only in the scale of the outburst (the light curves, for instance, are very similar in shape), their spectra are totally different. This fact scarcely leaves us in doubt that we are dealing with two classes of object which differ rather in kind than in degree, and which are probably radically dissimilar in their underlying processes.

The absence of hydrogen lines is likely to be of interest to the theoretical astrophysicist, who has the choice of attributing it either to very high excitation of ordinary stellar material or to normal excitation of material in which hydrogen is of low abundance. The history of spectroscopy should be enough to deter anybody from making a hasty correlation between atomic abundances and line intensities, but in this case such a correlation may be significant. Surface temperatures of several hundred thousand degrees, such as have been suggested [3] for supernovae, would strip the lighter atoms to their cores, and the energy would be mostly concentrated in the unobservable soft X-ray region. The observed distribution in the visual and photographic regions does not support this view. The low-abundance hypothesis, however, is not intrinsically unlikely ; we know now that the outburst of an ordinary nova blows the surface layers (consisting mostly of hydrogen) away from what is presumably an ordinary star which in its evolutionary path has reached a

* Astrophysicists using the word " band " in its spectroscopic sense do not necessarily (or even usually) imply that the emitter is a molecule. The reference is often, as seems probable in this case, to an atomic line broadened by Doppler effect.

critical mass and luminosity ; is it not possible that in the more violent eruption of a supernova much more matter is ejected, necessarily from the interior, where, as modern theories of stellar constitution [4] assure us, the hydrogen abundance is much less ?

The problem of identification is much complicated by the great widths of the lines. If the spectrum at maximum consists of partially overlapping emission bands without a continuous background, a temperature obtained from the observed energy distribution can be of no significance. Calculation of the general level of excitation from such temperatures can therefore be of no help in identifying the lines. Conversely, until the lines are positively identified we can have little idea of the state of excitation of the emitting matter, and are thus in the dark as to the physical and radiative structure of the surface responsible for the observed radiation. In other words, we cannot determine the energy distribution until the spectra are identified, and we cannot identify the spectra until we know the order of excitation, i.e. the temperature.

However, certain negative deductions enable us to avoid this difficulty, at least partially. For instance, Minkowski [1] has found that even on heavily over-exposed spectrograms of SN IC 4182, no trace of emission appeared below λ 3300. Since no identifications are available for the lines in other regions, there is no means of deciding whether emission lines are to be expected in the ultra-violet, but at any rate there is at present no reason why continuous absorption should be suspected there. The absence of even a trace of continuous spectrum does, then, suggest that the effective temperature is not outside the ordinary stellar range. This conclusion is strengthened by a similarity, as regards general appearance and widths (though not positions) of the emission bands, between the spectrum of a supernova a few weeks after maximum and that of a Wolf-Rayet star. If this similarity extends also to excitation level, we must expect in the spectra of supernovae lines whose excitation, though high, is not so high as to prevent us from knowing anything of the transitions involved.

Let us take now, in chronological order, observational details of supernovae spectra. Nothing at all is of course known of the appearance of the pre-outburst spectrum ; we cannot do more than guess at the nature of the object before the cataclysm, our only pointer here being theory [5], which indicates that collapse of a star on to a neutron core may be the result of a not unusual type of instability, and will lead to an energy output sufficient to account for a supernova.* For details of the spectrum just before maximum light we must rely on three observations of SN NGC 1003 made two days before maximum. These show a somewhat uneven continuous spectrum crossed by what may be wide absorption lines. Tentative assignments of these lines have been made [1] to He II and O IV. If these are correct, the lines are displaced about 45 A. towards the violet. This displacement is in general agreement with a half-width of 50 A. shown by the narrowest emission bands

* Readers may wish to note that this question is also alluded to in the *Report* by Bethe and Marshak, p. 1. [Editor.]

in somewhat later spectra, and indicates an expansion velocity of 3000 km./sec. Owing to the nature of the spectra, however, the evidence cannot be considered conclusive of the presence of these lines.

For a week or so after maximum the spectrum is so hazy that almost any line might be present and yet escape identification; at this stage it can be described only as an irregular continuous spectrum. After two or three weeks, however, the emission bands become more conspicuous, and individual maxima can be traced from day to day. The spectra of all supernovae show the same bands with remarkably similar widths, and there is now general agreement that the widths are, at any rate as a working hypothesis, to be interpreted as due to the Doppler effect of an atmosphere expanding, or rather exploding, with a velocity of the order of 6000 km./sec. This is some five times as great as in ordinary novae. At this stage Minkowski [1] has noted an extraordinary division of the spectrum, as regards behaviour, at $\lambda=5000$. At longer wave-lengths, pronounced changes occur from week to week, almost from day to day; bands appear and disappear much as in ordinary novae. In fact the general appearance is that of the spectrum of a source (or a succession of sources) undergoing rapid changes both in excitation and in radial velocity. Below $\lambda=5000$, however, a permanent pattern forms, consisting of a strong band at $\lambda=4575$ and weaker ones at 4300, 4800 and 4900 A.; and as the outburst dies down this pattern shifts towards the red, rapidly at first, more slowly later, by an amount which increases with wave-length, which is remarkably constant for different supernovae, and which in SN IC 4182 reached 125 A. for the band at $\lambda=4900$ a year after maximum. This behaviour is not accompanied by similar progressive displacements in the red bands, and seems at present almost inexplicable, particularly as no significant changes occur in the intensity distribution, as would be expected if the effect were merely a chance result of alterations in intensity of the various lines.

We can only conclude that the permanent bands below $\lambda=5000$ differ, certainly in nature of origin, and probably also in place of origin, from the others. This is confirmed by a survey of the intensity changes in the two regions. The red part strengthens at first, and at about 40 days after maximum becomes preponderant; in a further 120 days it fades to a minimum which is followed 20 days later by a sharp rise to almost its former intensity; finally it fades slowly. At no time have these intensity changes any parallel in the blue region; nor do the red bands appear to show the red shift so noticeable below 5000 A. in the later stages.

The strongest and most permanent bands beyond $\lambda=5000$ are at $\lambda=6500$ and $\lambda=5865$. Slight fluctuations of their wave-lengths with epoch are probably due to changes in composition of the bands rather than to any underlying physical cause such as is indicated by the red shift of the blue bands. It is significant that no bands in the blue region remain strong when the red portion of the spectrum becomes preponderant 40 days after maximum; this precludes any identification of the red bands with lines which would normally be accompanied by ones of shorter wave-length than 5000 A. This very much reduces our scope: for instance, Hα cannot be a prominent constituent of the band

at $\lambda=6500$, nor can He 5876 be mainly concerned in the emission of the band at $\lambda=5865$. The identification of $\lambda=5865$ with the D lines is negatived both by the general level of ionization believed to be operative and by the absence of other Na I lines and those of Ca I and Ca II which would be expected to accompany them. There is, however, a distinct possibility that two bands at 6085 and 5300 A. which emerged and disappeared simultaneously in SN IC 4182, are to be identified with the forbidden Ca V lines 6086 and 5309 A. respectively. If this is so, we may console ourselves for our failure to trace the origin of the other bands by attributing it to incompleteness of knowledge, soon to be rectified, of the spectra of highly ionized atoms.

The difficulty of interpreting the permanent bands below 5000 A. is increased by the progressive alteration in their wave-lengths. The wave-lengths adopted for these bands will depend on whether we assume this shift to be an increasing displacement towards longer wave-lengths or a decreasing one towards shorter wave-lengths, or first one and then the other. The early identification [6] of the strongest band at about $\lambda=4600$ with the N III lines near 4640 A. was no doubt suggested by analogy with the spectrum of an ordinary nova, but the coincidence of wave-lengths is only within experimental error for a couple of months. Such temporary coincidence must be regarded as fortuitous, particularly as no features above 5000 A. follow the intensity changes of the blue band. The fact is that until the red shift is satisfactorily explained, no identifications of the blue bands can be convincing.

In the final stages of the outburst, at about 180 days after maximum, two bands, only 40 A. wide, appear at 6299 and 6359 A. As the other bands in the red fade, these gradually grow in intensity; and for some months before the supernova has faded beyond spectrographic reach, they are the only observable features in the red. There seems little doubt that these are the forbidden "nebular" O I lines at 6300 and 6364 A. There is no trace at any time of the auroral line at 5577 A., so that the lines are presumably excited at extremely low pressure. The final stage of the supernova spectrum consists mainly of these two bands and the three strong ones in the blue, these now being displaced by 100 A. or more from their positions just after maximum.

The problem of this red shift is evidently fundamental to the spectroscopic study of supernovae. No such shift occurs in the spectra of ordinary novae. Minkowski [1] has shown that it begins at the same stage in all supernovae spectra so far recorded, and develops at the same rate; in fact, he suggests that the wave-lengths found for the blue bands will in future give an accurate idea of the stage of development of a newly-discovered supernova. No systematic change in the widths of the bands accompanies the shift, nor does their structure change significantly. This rules out any possibility of explaining the shift as the simultaneous fading of one band and development of another at a distance small compared with the width of the bands; such an effect would be altogether too improbable. If the displacement is to be explained as a Doppler effect, it cannot be the shift which accompanies the expansion of a gaseous shell enveloping the star, for in that case the width of an emission line should increase simultaneously with the shift by twice its

amount. To explain the shift as a Doppler effect at all we must assume that the centre of the stellar disc emits most of the light, and we are thus driven to the picture of a stellar envelope, heavily darkened towards the limb, expanding rapidly at the time of maximum light emission and less rapidly as the supernova develops. The "red shift" is thus a decreasing violet shift, and the "true" wave-lengths of the bands are those observed at the end of the outburst. Unfortunately for this explanation, the amount of limb darkening which has to be assumed is so high as to seem improbable. It has been suggested [1], however, that this assumption can be avoided if the bands are widened "negative absorption lines" produced when the number of atoms in the upper (metastable) level of a line is higher than that in the lower.

The alternative view of the red shift is that it is gravitational in origin; Zwicky [3] has suggested that it is due to the increasing potential at the surface of a collapsing star. This again leads to the conclusion [1] that the lines are really forbidden negative absorption lines, but this time the effect is literally a red shift, and the "true" wave-lengths are those observed in the early stages of the outburst. This explanation involves no more assumptions than are already needed in the "collapse" theory, which accounts for the known emission of energy (amounting to some 10^{50} ergs) by the sudden contraction of a star following the creation of a neutron interior. This theory is perhaps the more attractive in that it draws an analogy between novae and supernovae, the former being caused by gravitational collapse of a star into degenerate matter, the latter by a similar collapse on to a much more condensed core of neutrons.

Both hypotheses advanced to account for the red shift, however, leave unexplained the observed absence of displacements in the red part of the spectrum. We are driven to assume either that the source of this radiation is totally different from that below 5000 A., or that the shift is masked by the rapid changes of structure of the red bands, rapid changes which we know have no counterpart in the blue bands.

We may thus now present a tentative picture of the processes accompanying the outburst of a supernova. First of all there is a sudden release of a large amount of energy in the interior of a star of moderate mass and not abnormal luminosity. How this energy is produced is a matter of controversy: one hypothesis [3] identifies it with the potential energy released by the sudden contraction of the outer layers of a star on to a neutron core; another [7] attributes it to the collision of two stars (such collisions can occur in the central regions of galaxies from one to thirty per cent as frequently as supernovae appear); yet another [8] suggests that the high temperatures produced by a grazing collision may initiate nuclear transmutation processes of the type now believed [9] to account for the generation of energy in main-sequence stars, and may thus be the trigger which sets off a sort of infectious collapse. The exact process involved is evidently by no means certain yet, and is anyway somewhat beside the point; it is quite possible that almost any sudden release of energy to the amount of 10^{50} ergs, however caused, would produce similar results. That the release is sudden is proved by the light curves: the average

supernova probably increases its brightness by a factor of the order of 10^8 in a matter of days. The light maximum is quite sharp (it can be located to within two or three days), and is followed by a fall of brightness, rapid at first but much less so after a few weeks, as in an ordinary nova. There is little evidence, however, for the rapid post-maximum fluctuations shown by many common novae. The general impression given by the light curves is that of the instantaneous release of a large amount of energy, followed by the gradual relaxation of the star into its new state. The total amount of energy radiated is believed [10] to destroy an insignificant fraction of the total mass—perhaps [11] one or two per cent.

The absorption spectrum just before maximum suggests the rapid expansion of the star's chromosphere, and already shows evidence of the expulsion of fairly highly ionized matter at velocities of a few thousands of kilometres per second. After maximum the spectrum changes to one showing emission bands whose widths imply a velocity of expulsion which appears to be fairly constant at 6000 km./sec. for all supernovae. The ejection process appears to be uniform, as in the case of Wolf-Rayet stars, rather than spasmodic, as in ordinary novae, from which shells of gaseous matter are often thrown with varying velocities at irregular intervals. The absence of hydrogen lines at this stage may be attributed either to high ionization or to low abundance of hydrogen in the emitting matter.

The observed changes as the outburst dies down, viz. from a series of overlapping emission bands to a few strong individual ones, are not unlike those observed in common novae, and if similarly interpreted imply a general rise in the state of excitation. Until the bands observed at this stage are positively identified, it will be difficult to decide whether they are more likely to be forbidden or permitted transitions. If the former, it will be tempting to describe the supernova as simply an ordinary nova writ large: the chief source of radiation will be the atomic processes induced in the low-density ejected matter by the high-temperature diluted radiation of the central core. If the latter, the analogy should rather be drawn with Wolf-Rayet stars: most of the lines originate close to the photosphere where the density is not excessively low and where continuous ejection is taking place. Meanwhile, as the star collapses into neutron matter and gives up its potential energy as high-temperature radiation (which does not, however, escape unchanged through the atmosphere, as is evidenced by the absence of ultra-violet continuum), the gravitational potential increases to enormous values at its surface, and the emission bands are shifted to the red by amounts of the order of 100 A.

At the final stages of the outburst the forbidden O I "nebular" lines become prominent, unaccompanied, curiously enough, by any of the other lines usually associated with them. Whether this is due to some peculiarity of excitation or to an increased abundance of oxygen as the result of nuclear reactions remains to be decided. It may be that these lines are the forerunners of the nebular spectrum which ordinary novae exhibit a few months after their outburst; if so, supernovae must be regarded as developing along the same lines as common novae, but much more slowly. This is after all only to be expected:

after an outburst of energy so much greater, the time of relaxation is naturally longer. That their ends are similar is almost certain : it is now accepted [12] that the Crab nebula is the remnant of a supernova which appeared in A.D. 1054, and it shows the ordinary nebular spectrum, indicating a velocity of expansion (1300 km./sec.) of the same order as that observed in common novae. It is therefore probably a giant example of the type of nebulous envelope surrounding such ordinary novae as N. Persei (1901) and N. Aquilae (1918).

If there are gaps and inconsistencies in this sketch of the phenomena attendant on a supernova outburst, we must remember that both theoretically and observationally the subject is as yet in its infancy. Parallel advances, made by the physicist in his knowledge of nuclear processes, by the astronomer in his knowledge of the conditions of stellar interiors, by the spectroscopist in his interpretation and prediction of the spectra of highly ionized atoms, and above all by the observer in obtaining spectra of yet undiscovered supernovae, will, no doubt, revolutionize our notions in the next decade. The best stimulant to such advances would be the appearance of a supernova in one of the local group of galaxies, or better still in our own ; with the wealth of material we should then obtain we could hardly fail to clarify our ideas on the subject.

REFERENCES

(1) Minkowski, R. *Astrophys. J.* **89**, 156 (1939).
(2) Strohmeier, W. *Z. Astrophys.* **14**, 227 (1937).
(3) Zwicky, F. *Publ. Astr. Soc. Pacific*, **50**, 215 (1938).
(4) Strömgren, B. *Astrophys. J.* **87**, 520 (1938).
(5) Baade, W. and Zwicky, F. *Proc. Nat. Acad. Sci., Wash.*, **20**, 259 (1934).
(6) Payne-Gaposchkin, C. *Astrophys. J.* **83**, 173 & 245 (1936).
(7) Whipple, F. L. *Proc. Nat. Acad. Sci., Wash.*, **25**, 118 (1939).
(8) Whipple, F. L. *Proc. Amer. Phil. Soc.* **81**, 253 (1939).
(9) Bethe, H. A. *Phys. Rev.* **55**, 434 (1939).
(10) Payne-Gaposchkin, C. *Proc. Nat. Acad. Sci., Wash.*, **22**, 334 (1936).
(11) Landau, L. *Nature, Lond.*, **141**, 333 (1938).
(12) Baade, W. *Astrophys. J.* **88**, 285 (1938).

II. SOME RECENT DEVELOPMENTS IN THE ZEEMAN EFFECT IN ATOMIC SPECTRA

By HARVEY E. WHITE, A.B., PH.D.
University of California, Berkeley, Calif.

§1. INTRODUCTION

SINCE the time that Zeeman first discovered the effect of a magnetic field upon spectrum lines, the *Zeeman effect*, as it is called, has grown steadily in importance until today it stands as one of the high-lights of modern scientific achievement. To the theoretical physicist the Zeeman effect forms an integral part of the modern quantum theory of matter and radiation, while to the experimental physicist it is a means by which the analysis of a given spectrum is made possible.

It is in the latter case—that is, the experimental field of investigation and in the analysis of spectra—that the most recent progress has been made. This statement is not intended to imply that the theory is not well developed, for, on the contrary, the theory of the Zeeman effect has been so well formulated that practically all existing experimental data are well accounted for.

From the experimental standpoint, progress with the Zeeman effect has been made along two well-developed lines. There has been on the one hand the development of new and better sources of light, and on the other the construction of magnets capable of producing stronger fields and uniform fields of larger dimensions. Although important contributions in this direction have been made by many different experimenters in many different laboratories all over the world, only four particular developments have been selected for detailed presentation in this *Report*.

§2. STRONGER SUSTAINED MAGNETIC FIELDS

The importance of Zeeman-effect data as an aid to the analysis of an uncharted spectrum has long been recognized by every spectroscopist. A full realization of this can be obtained by anyone who has ever observed or photographed the spectrum of an element like iron, or titanium, or one of the rare earths like cerium or praseodymium. For any one of these elements the visible arc spectrum alone contains thousands of spectrum lines arranged with little or no apparent regularity.

The value of a completely resolved Zeeman pattern for any one spectrum line is recognized when it is recalled that such a pattern enables one to determine, among other things, the important quantum numbers of both energy levels involved in the transition that gives rise to that line. It is principally for this reason that repeated attempts have been made by numerous investigators to produce stronger and stronger magnetic fields. In general the stronger the magnetic field the wider is the separation between the adjacent

Zeeman components of a spectrum line and the easier it is to resolve and measure the resulting line pattern.

Until recently sustained field strengths as high as 43 kilogauss have been produced in a number of laboratories where the proper equipment, usually a Weiss magnet, has been permanently installed. This apparent limiting value, which is determined by the saturation of the iron used as a core in the electromagnet, has been exceeded recently by Jacquinot [1] and his colleagues using the large Bellevue magnet in France. Field strengths as high as 65·8 kilogauss have been attained by forcing a larger current through the coils than that necessary to saturate the iron.

A careful study of the theory of electromagnets by F. Bitter [2] has recently led to his development of several new magnets, the latest one of which is an air-core type capable of producing sustained field strengths up to 100 kilogauss. The coil of this magnet, enclosed in the cylindrical bronze jacket shown at the centre in figure 1, is composed of 200 turns of copper ribbon through which currents up to 10,000 amp. can be sent continuously. Each of the copper bus bars leading the current to and away from the magnet consists of a bank of eight 8-inch by $\frac{1}{4}$-inch copper strips. Power is furnished at 170 v. by a specially-designed 1700-kw. motor-generator set. A manual control of the current through the magnet keeps the field strength constant to within 0·1 per cent. Heat is removed by water which, flowing to and from the coil through 4-inch pipes, passes at the rate of 800 gallons per minute.

The field at the centre of the coil is sensibly constant over a volume of about one cubic inch. Such a large volume as this has made possible the design, by G. R. Harrison [3], of a new type of arc source. Two metal electrodes mounted coaxially in a tube-like holder enable the arc stream to be lined up with the field lines. This is a necessary condition to keep the arc burning. A spare arc with water-cooled electrodes is shown lying on the shelf at the right in figure 1. To conserve space both electrodes are offset from the axis of the tube, and a small quartz right-angled prism is mounted just beside the electrodes to reflect the transverse light from the arc out through the axis of the coil to the slits of three spectrographs. Most important is the use of an air-jet blown between the arc and the prism to keep the prism face nearest the arc cool and clean.

Arc electrodes $\frac{1}{8}$ inch in diameter, made of silver impregnated with salts of the element to be studied, are burned steadily with a current of about 4 amp. Sending the light to the slits of three 35-ft. grating spectrographs simultaneously enables high dispersion spectrograms covering the range 2000 to 8000 A. to be made with each exposure. Up to the present time complete sets of plates have been obtained for columbium (niobium), iron, rhodium, ruthenium, thorium, thulium, tungsten, cerium, erbium, europium, gadolinium, neodymium, praseodymium, and ytterbium.

On most plates, lines of the singly and doubly ionized atoms are more in evidence than those of the neutral atoms. A typical portion of a rhodium plate is reproduced in figure 2, while enlargements of selected patterns in the rare-earth spectra of europium and neodymium are given in figures 3 and 4.

Figure 1. Photograph of the powerful electromagnet developed at the Massachusetts Institute of Technology by F. Bitter.

[*To face p.* 146

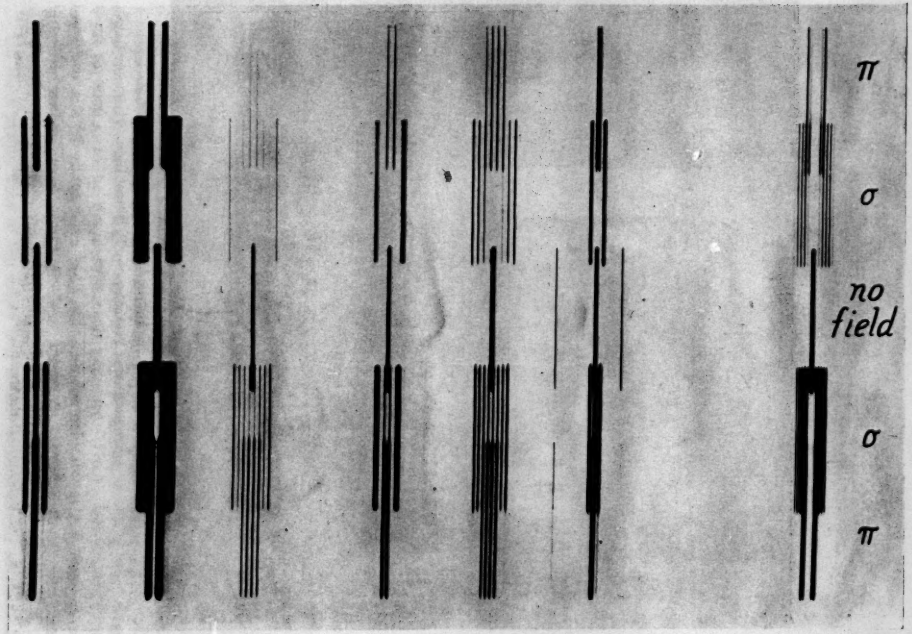


Figure 2. Portion of the spectrum of rhodium in the region 3500 to 3460 A. The π and σ components at the top were taken with a field strength of 90·5 kilogauss, and those at the bottom with a field strength of 70 kilogauss. (After Harrison and Bitter.)

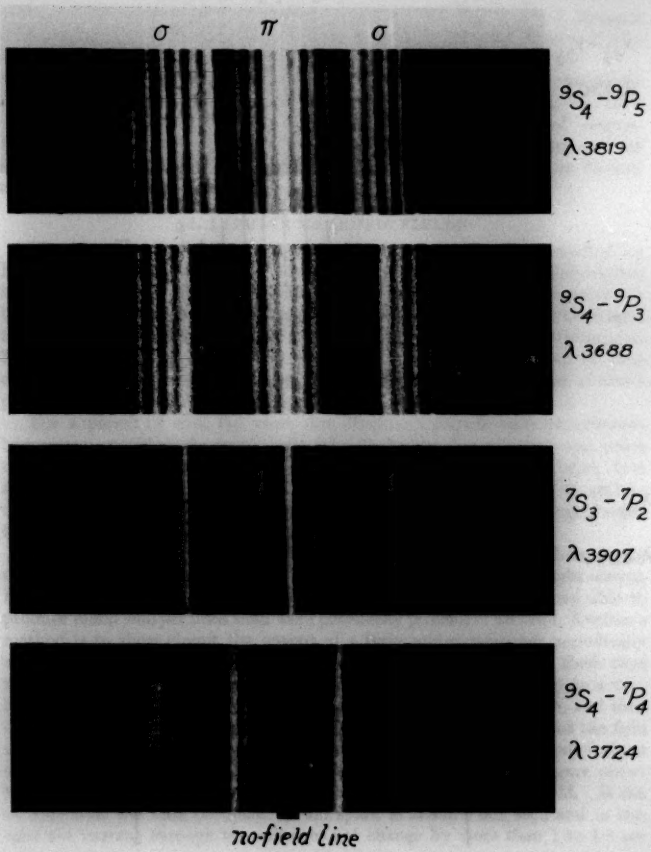


Figure 3. Zeeman patterns of four lines in the spectrum of singly-ionized europium, Eu II. Field strength 85 kilogauss. (After Harrison and Albertson.)

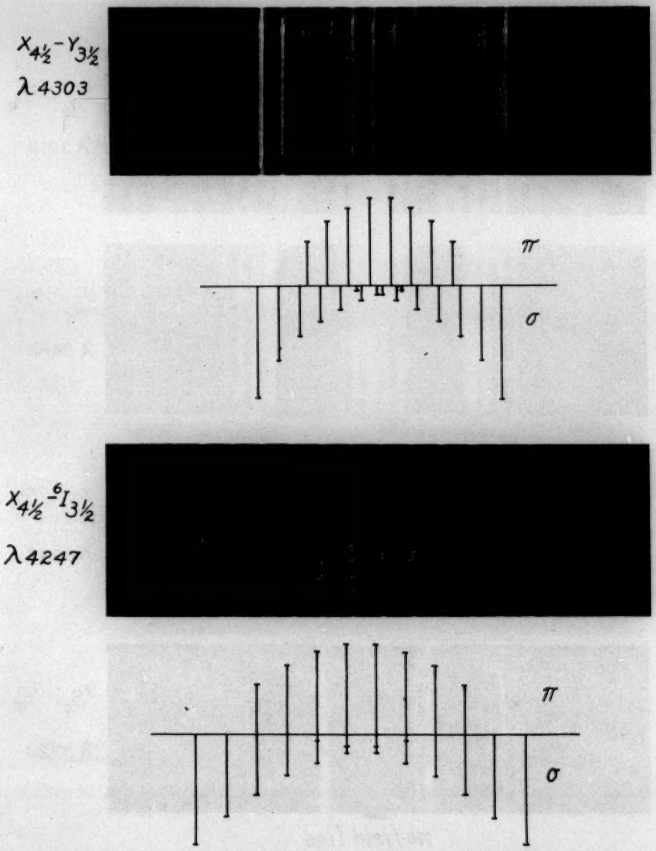


Figure 4. Zeeman patterns of two lines in the spectrum of neodymium, Nd. Field strength 96 kilogauss. (After Harrison and Albertson.)

[*To face p.* 147

Were it not for Zeeman patterns such as these, the analysis of most of the elements listed in the preceding paragraph would be practically impossible; but in the hands of such experts as H. N. Russell, W. Albertson, W. F. Meggers and others, the last and most complicated spectra are being methodically analyzed.

Only recently, for example, W. Albertson [4], with the help of completely resolved Zeeman patterns of many lines photographed by Harrison, has been able to analyze the spectrum of ionized europium and of ionized cerium. It is interesting to point out that the spectrum of ionized europium exhibits septet and nonet terms, while that of the neutral atom, analyzed by Russell and King [5], contains octet and decet terms.

§ 3. IMPULSE MAGNETIC FIELDS

The Zeeman effect in very strong magnetic fields was first studied by Kapitza and Skinner [6] in 1925, the fields being produced by a method developed in the previous year by Kapitza himself. By short-circuiting an accumulator battery through a heavy copper coil, devoid of iron, they were able to obtain, for periods of about a hundredth of a second, field strengths up to 140 kilogauss. The Zeeman patterns obtained in these early experiments were not all that could be desired, the chief criticism being that the component lines of every pattern were too broad and diffuse.

For a period of over ten years now Kapitza's impulse-field experiments have been continuously improved. The technique has reached the point where well-defined and sharp patterns of lines can be produced with field strengths up to 300 kilogauss. The first improvement made was to replace the battery by a motor-generator which, when short-circuited, gave greater field strengths, uniform over considerably larger volumes.

The most recent improvement, and apparently the most important, is the reduction of the current density at the spark discharge used as a light source. In making this change, Kapitza, Strelkov and Laurman [6] have been able to produce much sharper lines than were previously possible. In brief, Kapitza's method is to short-circuit the output of a large motor-generator periodically with a coil containing a number of turns of heavy copper wire. Each time the switch is momentarily closed the current in the coil rises quickly to a very high maximum, as shown by the oscillogram reproduced in figure 5, and then drops off, with equal rapidity, to zero. Owing to this impulse current the field rises in exactly the same way and falls to zero again in a time interval of about 0·03 sec. A second oscillograph trace across the centre of the figure shows the time interval of the spark occurring at the peak value of the field. It can be seen that the time occupied by the spark is about 0·001 sec., and in this time the current through the coil does not change by more than 1 to 1·5 per cent.

A series of five photographs of the Be II doublet at 3131 A. is reproduced in figure 6. The field strengths employed for these exposures were 0, 74, 168, 241 and 307 kilogauss respectively. Each picture, showing σ-components only, is in excellent agreement with theory at every point. Not only are the

separations of the components correct, but the relative intensities as well. The normal Zeeman triplet separation drawn just below the last photograph is given for comparison with theory. A more detailed agreement is shown in the graph at the right in figure 6. In brief, the theory for the Paschen-Back effect for a *principal-series doublet* shows that with increasing field strength the symmetrical patterns of each line of the doublet spread out, cross each other, and then form approximately a *normal triplet*. Each σ-component of the triplet should be a doublet, symmetrical about the normal triplet position, with two-thirds the separation of the original no-field doublet. Furthermore, the other two σ-components should approach twice the normal triplet separation and at the same time gradually fade out, as observed. The heavy lines in the graph represent theory and the dots represent experimental measurements taken from the photographs.

§ 4. THE ZEEMAN EFFECT OF HYPERFINE STRUCTURE

An extension of the Zeeman effect to the study of hyperfine structure of spectrum lines was first made by E. Back and S. A. Goudsmit [7] in 1928. The published report of their observations is now considered a classic in the annals of atomic spectra. Although similar observations have since been extended to several other elements, the methods employed have been confined to heavy elements like caesium and thallium where the line widths are small and the structure broad enough to permit suitable resolution with available spectrographs.

As is nearly always the case in a study of hyperfine structure, there are two principal factors limiting the detail of the structure of the patterns: (1) the sharpness of the spectrum lines, and (2) the resolving power of the spectroscopic instruments.

An admirable contribution to this phase of spectroscopic research has recently been made and published by D. A. Jackson and H. Kuhn [8]. Instead of lowering the effective temperature of their light source to reduce Doppler broadening of the lines, they have developed an entirely new process known as the *molecular beam* method of observation. The molecular beam is produced in a tube of special design as illustrated in cross-section in figure 7. Sodium metal at the bottom of the tube is heated to a point where sodium atoms are evaporated rapidly from the surface. As the atoms fly upward through the highly evacuated tube, there are relatively few collisions between them, and each atom follows a straight-line trajectory to the top, where it sticks to the glass wall.

A parallel beam of light from a sodium arc, emitting the well-known yellow lines at 5890 and 5896 A., is then made to pass through the atomic beam, after which the light is focussed on the slit of a prism spectrograph. Since the motion of the absorbing sodium atoms is at right angles to the incident light beam, the Doppler broadening is extremely small. With a tube ten times as long as it is wide the maximum component of velocity in the direction of the light beam is $\frac{1}{10}$ the full velocity. This Jackson and Kuhn refer to as a 1/10 collimation of the beam, and by the theory of the Doppler effect should give

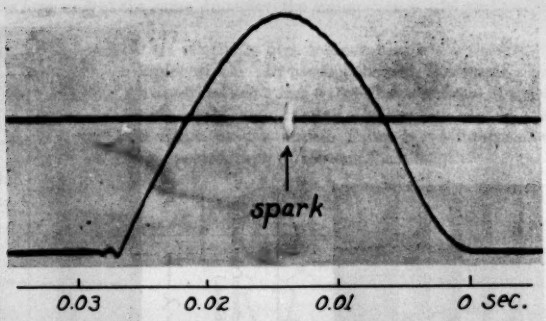


Figure 5. Oscillograph curves showing the spark discharge at the time of maximum field strength. (After Kapitza, Strelkov and Laurman.)

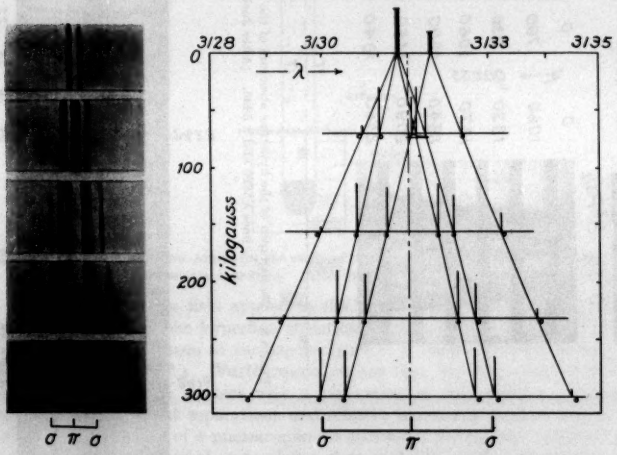


Figure 6. Paschen-Back effect of the Be II doublet at λ3131, photographed at field strengths up to 300 kilogauss.

[To face p. 148

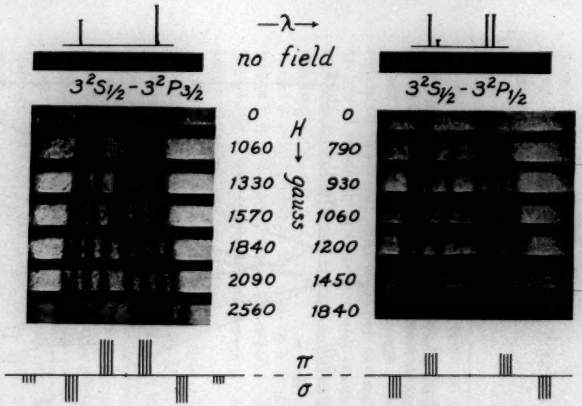


Figure 8. Photographs of the hyperfine structure of the Zeeman effect of the sodium yellow lines λ 5890 and λ 5896. (After Jackson and Kuhn.)

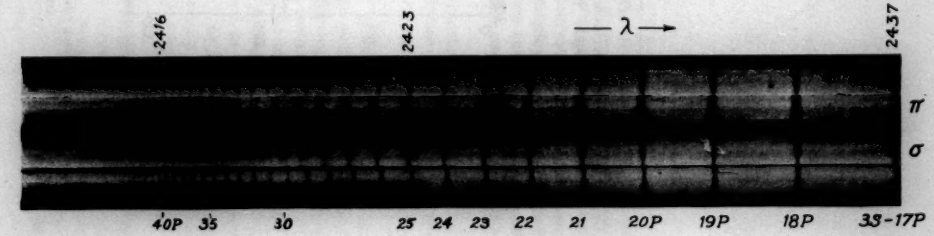


Figure 11. Photographic reproduction of the quadratic Zeeman effect as observed in the principal series of sodium. (After Jenkins and Segrè.)

a half-value breadth to each spectrum line of only 0·006 cm.$^{-1}$ With a tube twice as long and of the same diameter, the collimation is 1/20 and the Doppler half-value width is 0·003 cm.$^{-1}$

The arc source with its relatively high temperature emits the yellow lines with a considerable Doppler width. In passing through the atomic beam the centres of these lines are absorbed according to the characteristics of the sodium atoms. To obtain a sufficiently high resolving power for the detailed line structure, Jackson and Kuhn used a double Fabry-Perot interferometer in conjunction with a prism spectrograph of 150 cm. focal length.

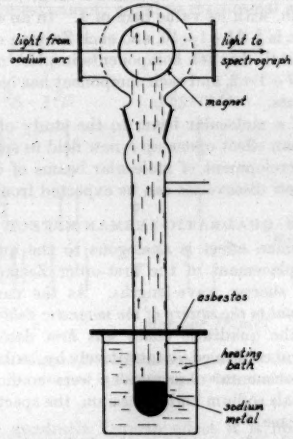


Figure 7. Molecular-beam source for the study of the hyperfine structure of the Zeeman components in sodium. (After Jackson and Kuhn.)

Without a magnetic field applied to the beam, and with a collimation of 1/20, photographs of the hyperfine structure of each sodium line have been made. Each line, as seen at the top in figure 8, is itself a doublet having a separation of only 0·02 A. Furthermore, on one line, $^2S_{\frac{1}{2}}-^2P_{\frac{1}{2}}$, each of these components is itself a doublet with a separation of only 0·002 A. Careful measures of component separations and relative intensities confirm previous assignments by others of a nuclear spin for sodium of $I=\frac{3}{2}(h/2\pi)$. According to these results the normal state of the atom is double with a separation of 0·0592 cm.$^{-1}$, the first excited level, $3\,^2P_{\frac{1}{2}}$, is double with a separation of 0·0064 cm.$^{-1}$, while the companion level, $3\,^2P_{\frac{1}{2}}$, is quadruple, but so narrow that it cannot be resolved.

Using a collimation of 1/10 and applying weak magnetic fields to the atomic beam as shown by the magnet in figure 7, Jackson and Kuhn obtained the well-resolved Zeeman patterns with hyperfine structure reproduced in figure 8.

Although these pictures show only the π-components of both lines, the positions of the hyperfine components shift exactly as predicted by theory for the corresponding field strengths. Presumably the σ-components would agree equally well with theory. Note how in very weak fields the Zeeman components group into sets of 5 and 3, whereas in strong fields one component shifts over with the three to produce symmetry with two groups of four components each.

Each set of four lines represents only the π-components of the regular Zeeman pattern for a principal series doublet as shown at the bottom of the figure. From theory the number 4 is obtained from $2I+1$, where I is characteristic of the nuclear spin, with its value here of $\frac{3}{2}$. In an element like bismuth, where the nuclear spin is $\frac{9}{2}$, $2I+1=10$, and each Zeeman component, as shown by Back and Goudsmit [7], has ten components in place of four. In thallium, where the spin $I=\frac{1}{2}$, $2I+1=2$, and each component has been found by Wulff [9] to have two components.

The application of a molecular beam to the study of hyperfine structure as well as to the Zeeman effect opens up a new field in spectroscopic research. For the future, the development of molecular beams of other elements looks very promising, and new discoveries can be expected from it.

§ 5. QUADRATIC ZEEMAN EFFECT

The quadratic Zeeman effect is analogous to the quadratic Stark effect and consists of a displacement of the first-order Zeeman components of a spectrum line toward shorter wave-lengths. As the name implies, this displacement is *proportional to the square of the magnetic field-strength*.

The existence of the quadratic effect was first detected experimentally by Segrè [10] in 1934, and confirmed quantitatively by Jenkins and Segrè [11] five years later. The experimental observations were confined to the principal series of the alkali metals sodium and potassium, the spectrum being produced in absorption in each case.

In both the classical and quantum-mechanical treatments of the first-order Zeeman effect the theory is generally simplified by leaving out a second-order term in the energy equation. The reason for omitting this term is that the calculated effect is relatively small and is in general not observable.

The classical theory of the Zeeman effect, as first treated by H. A. Lorentz and later extended by quantum mechanics, gives for the quadratic Zeeman effect

$$\Delta\nu_Q = \frac{e^2H^2}{8mc^3h}\, a_1 n^{*4}(1+m_i^2), \qquad \ldots.(1)$$

where a_1 is the radius of the first Bohr circular orbit and equal to $h^2/4\pi^2e^2m$, n^* is the *effective quantum number* of the upper energy level, and m_i is the orbital magnetic quantum number equal to ± 1 for the σ-components and 0 for the π-components.

By substituting the most probable values of the physical constants in this equation the quadratic formula becomes

$$\Delta\nu_Q = 4{\cdot}98\times 10^{-15} n^{*4}(1+m_i^2)\ \text{cm}^{-1} \qquad \ldots.(2)$$

When this quadratic term is added to the normal first-order Zeeman effect,

$$\Delta\nu_N = 4{\cdot}67 \times 10^{-5}\, H \text{ cm.}^{-1}, \qquad \ldots.(3)$$

one finds that the two δ-components of each normal triplet preserve their separation $2\Delta\nu_N$, but that their centre of gravity (see diagram (c) in figure 9) is shifted to shorter wave-lengths by an amount twice that of the π-components. It will be seen from the relative magnitudes of the two constants in equations (2) and (3) that for small n^* and H the quadratic effect is negligibly small. For a field strength of 27 kilogauss, however, and a value of $n^* = 25$ the quadratic shift $\Delta\nu_Q$ becomes approximately equal to the normal splitting $\Delta\nu_N$.

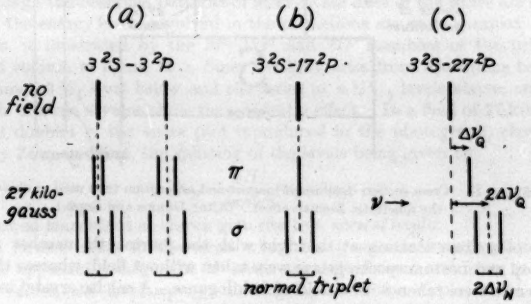


Figure 9. Schematic diagram showing the development of the quadratic Zeeman effect as observed in the principal series of sodium.

To observe the quadratic Zeeman effect it is necessary therefore to go to spectral lines involving large n^* values. The principal series of sodium and of potassium were used by Jenkins and Segrè, since in absorption the lines can be observed as far as $n = 40$ or 50.

The experiment was made possible through the use of the new cyclotron magnet of the Crocker Radiation Laboratory at the University of California. The pole faces of the 225-ton magnet (figure 10), with a diameter of 152 cm. and a separation of 56 cm., lend themselves admirably to the use of a long absorption tube. The gap was reduced to 4·4 cm. by the insertion of extra iron blocks, the final pole faces being 15 × 63 cm.

The absorption tube was of Pyrex tubing 3·5 cm. in diameter and 110 cm. long. The central portion, 30 cm. in all, was heated by nichrome wire to a temperature of 500° C. After the tube had been thoroughly evacuated, and small lumps of the alkali metal placed along the heated part of the tube were fused, pure hydrogen was introduced at about 10 cm. pressure to prevent diffusion of the metal to the windows.

A hydrogen lamp was used as a source for the continuous radiation, and

a 3-m. Littrow instrument containing two quartz prisms automatically adjusted to minimum deviation was used to photograph the spectrum.

The one set of photographs reproduced in figure 11 shows the principal

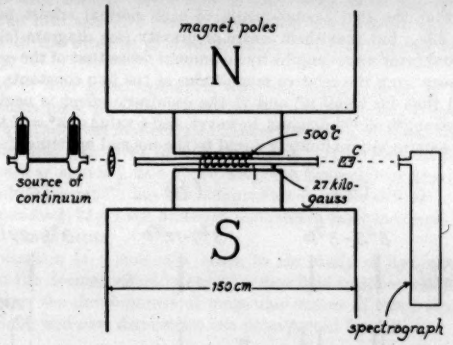


Figure 10. Cross-section diagram of magnet and absorption tube used in observing the quadratic Zeeman effect. (After Jenkins and Segrè.)

series of sodium starting at the right with the fourteenth member $3S-17P$. The top and bottom spectrograms were taken without field, whereas the two at the centre were taken with a field of 27 kilogauss. A calcite crystal in the path

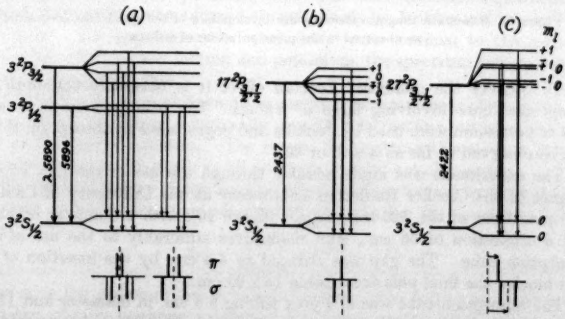


Figure 12. Zeeman splitting of the energy levels in sodium, illustrating the quadratic shift of levels with large principal quantum numbers.

of the light enabled the π- and σ-components to be photographed separately. It should be noted that starting at the above member of the series only the *first-order Zeeman effect* is observed, i.e., taking the π- and σ-components together,

the pattern of each line is a *normal triplet*, symmetrical about the no-field line. As one proceeds to higher members of the series the quadratic Zeeman effect appears, the π- and σ-components shifting more and more to shorter wave-lengths.

As a result of perturbations from neighbouring energy levels of the atom, the lines begin to broaden at about $n=28$, and at $n=32$ have practically broadened into a continuous absorption. It should be noted how much farther out in the series the no-field lines are observed. The cause of the broadening has been investigated thoroughly and a theoretical treatment has been given by Schiff and Snyder [12], but because of its complexity this will not be presented here.

Although the observed patterns of most of the lines of the series are simple triplets, the energy levels involved in the transitions are more complex. This situation is illustrated by the $3P$, $17P$ and $27P$ members of the principal series of sodium in figure 12. Since the lines arise from transitions between the common $3\,^2S_{\frac{1}{2}}$ level below and the series of $n\,^2P_{\frac{1}{2},\frac{3}{2}}$ levels above, only the ^2P levels of large n value show the quadratic effect. In a field of 27 kilogauss the first doublet of the series (not reproduced in the photograph) shows the ordinary Zeeman effect, the splitting of the levels being given by

$$\Delta\nu_N = \frac{eH}{4\pi mc^2}(m_l + 2m_s) \qquad \ldots\ldots(4)$$

The allowed transitions as shown give rise to a *normal triplet*.

Still farther along in the series where the quadratic term, equation (2), becomes appreciably large, the ^2P levels are shifted more and more to higher energy values, the level $m_l=\pm 1$ undergoing twice the displacement of the levels $m_l=0$. This is illustrated for the 24th member of the series, $3S-27P$, in diagram (*c*).

Accurate measurements of the displacements of both the π-and σ-components show excellent agreement with theory, following the quadratic formula as far as $n=21$. From there on, the shifts increase more rapidly than n^4 and begin to show other anomalies. At about this same point the lines, particularly the π-components, begin to show wings on the long wave-length side. Microphotometer traces of the original spectrograms indicate that the wing in each case is actually a faint, nearly unresolved line that increases in intensity until at $n=30$ it is practically equal to the parent line.

Both of the above discrepancies are explainable as follows. The elementary theory considers the ^2P terms alone and neglects the perturbing influences of nearby energy levels of the same parity, but different L-value. In particular the ^2F levels become so close to the ^2P levels that an appreciable "mixing" of the levels exists. Because the magnetic splitting of the levels becomes comparable to the distance between ^2P and ^2F terms, the eigen-functions of the ^2F states become mixed with those of the ^2P states and there is a mutual repulsion of the respective levels accompanied by forbidden transitions between the $3\,^2S$ normal state and the $n\,^2F$ states. These forbidden lines are the long-wave-length components or "wings" on the lines mentioned above. The fact that the lines do not appear in potassium further confirms this explanation.

In view of the recent developments presented here of new techniques for observing the Zeeman effect, spectroscopy should become more and more interesting as a field of experimental research. There is a particular satisfaction to be obtained from the knowledge that at last suitable methods are now available for analyzing those extremely complex spectra known to arise from the heavy elements in the centre and lower right-hand side of the atomic table. With such fine Zeeman-effect data available it becomes a pleasure to do what at one time appeared to be a hopeless task, namely, to analyze the known spectra of the rare-earth elements.

REFERENCES

(1) Jacquinot, P. and Belling, T. *C.R. Acad. Sci., Paris*, **201**, 778 (1935).
(2) Bitter, F. *Rev. Sci. Instrum.* **10**, 373 (1939).
(3) Harrison, G. R. and Bitter, F. *Phys. Rev.* **55**, 1125 (1939).
(4) Albertson, W. and Harrison, G. R. *Phys. Rev.* **52**, 1209 (1937).
(5) Kapitza, P. and Skinner, H. W. B. *Proc. Roy. Soc.* A, **109**, 224 (1925).
(6) Kapitza, P., Strelkov, P. G. and Laurman, E. *Proc. Roy. Soc.* A, **167**, 1 (1938).
(7) Back, E. and Goudsmit, S. A. *Z. Phys.* **47**, 174 (1928).
(8) Jackson, D. A. and Kuhn, H. *Proc. Roy. Soc.* A, **167**, 205 (1938).
(9) Wulff, J. *Z. Phys.* **69**, 74 (1931).
(10) Segrè, E. *Nuovo Cim.* **11**, 304 (1934).
(11) Jenkins, F. A. and Segrè, E. *Phys. Rev.* **55**, 52 (1939).
(12) Schiff, L. I. and Snyder, H. *Phys. Rev.* **55**, 59 (1939).

III. THE MAGNETIC EFFECT IN DIATOMIC SPECTRA

By F. H. CRAWFORD, Ph.D., Read Professor of Physics,
Williams College, Williamstown, Mass.

§ 1. INTRODUCTION AND HISTORICAL RÉSUMÉ

The now famous observations made by Zeeman in 1896 on the effect of the magnetic field on the sodium D-lines served as the starting point for a series of theoretical and experimental discoveries that was to continue for a generation. It is perhaps fortunate that the first line patterns actually resolved were simple triplets, which were almost immediately described by Lorentz in terms of the electron theory. In fact this now familiar relation for the " half-width ", $\Delta\nu_h$, of these triplets,

$$\Delta\nu_h = \frac{e}{4\pi m_e c} H = 4{\cdot}674 \times 10^{-5}\, H \text{ cm}^{-1}, \qquad \ldots.(1)$$

gave directly the ratio of charge to mass, e/m_e, of the new particles, the sign of which came from the polarization of the pattern components. Here, as throughout, c is the velocity of light *in vacuo*, H the applied external field in gauss, and energy changes will be reduced to their equivalent values in inverse centimetres by division by hc *.

This correlation of theory and experiment was so universally accepted that, although more complicated patterns not capable of interpretation on this basis were reported only a year later by Zeeman himself, and by Cornu, Preston and Michelson, these were regarded simply as " anomalous ", and attention was focussed on the symmetry of the effect, the study of the polarization of the components and, ultimately, its use in the classification of lines in series.

In the case of the magnetic effect on the lines of *molecular* spectra, which for conciseness we may refer to simply as the Zeeman effect in band spectra, no such elegant simplicity was observed. In fact, owing to the (by no means unformidable) experimental difficulties, the results for the first decade were entirely negative [3]. When, with the necessary improvements in technique and spectroscopic equipment, noticeable results were recorded they were so complicated, so varied and so different from those expected (from atomic analogies) as to be of small help in the development of band theory or the systematizing of band types.†

Actually Dufour [6], [46], [47] first succeeded in detecting definite effects in

* The notation throughout is essentially that of Mulliken " On the Interpretation of Band Spectra " [1] and of Jevons' *Report* [2] (Appendix I, pp. 262–265).

† For a more detailed account of some of the earlier experimental work reference is made to résumés, each fairly complete at its time of publication, by Kemble [4], Jevons [2] (chapter VII) and Crawford [5].

the (scarcely resolved) $^2\Pi \rightarrow {}^2\Sigma$ bands of CaF, CaCl, SrF, SrCl, BaF, BaCl and SiF and in certain lines of the many-lined spectrum of H_2. Results of analogous nature were next reported on the band heads of YO by Becquerel [7], on heads of certain bands of the $^3\Pi \rightarrow {}^3\Sigma$ first positive bands of N_2 by Croze [8] (who found negative results in the $^3\Pi \rightarrow {}^3\Pi$ second positive bands of N_2 and the $^2\Sigma \rightarrow {}^2\Sigma$ bands of N_2^+, but was able to extend Dufour's work on the H_2 lines). There were very extended investigations by Fortrat [9], [23], [38], [40], [44], by Deslandres and d'Azambuja [10], [37] by Deslandres and Burson [11] and Hulthén [12] on bands of a number of systems, including the $^2\Sigma \rightarrow {}^2\Sigma$ bands of N_2^+ and CN, the $^2\Sigma \rightarrow {}^3\Pi$ NO (γ), ZnH, CdH, HgH, OH λ3064 and CH λ3900 bands, the $^3\Delta \rightarrow {}^3\Pi$ CH band at λ4390, the $^3\Pi \rightarrow {}^3\Pi$ second positive of N_2 and Swan bands of C_2. We see at once that these bands, selected primarily for their ease of excitation, include some of the most complicated electronic types (with up to 26 branches actually present), and since they were studied in large part even before Fortrat [13] had recognized the importance of band-origins in band theory, they could hardly give results of readily generalized nature. Important empirical results (which are in harmony with later theoretical predictions) may in part be summarized as follows:

(1) The Zeeman patterns of band lines are often of the same size as those obtained in atomic spectra but are usually smaller, do not always change linearly with the field and are (apparently) only incompletely resolved in the vast majority of cases.

(2) The patterns vary from line to line in a branch and are different for the same line in different branches. The *earlier* lines in a given branch are usually most sensitive to the field, the effects observed decreasing rapidly with J'', the final quantum number of the line involved.

(3) The patterns exhibit polarizations of outer components which are usually similar to, but often just the reverse of, those in atomic spectra. Thus, in observation at right angles to the field (to which this discussion is confined) the pattern in $\perp$ polarization * is usually wider than that in $\parallel$ polarization, though the reverse also occurs.

(4) The patterns exhibit asymmetry in both *position* and *intensity* of components, the asymmetry increasing rapidly with the field.

(5) The narrow doublet (and triplet) members of multiplet branches are drawn gradually together at higher fields, with, at still higher fields, a single broad band filling roughly a region of the width of the original multiplet.

(6) In the case of the alkaline-earth halide bands (all of type $^2\Pi \rightarrow {}^2\Sigma$), a distinct doubling of the band heads is observed in each polarization, the average displacement being about 90 per cent of the normal triplet width of $2\Delta\nu_n$.

On the theoretical side, meanwhile, Lenz [14] (on a rather insecure theoretical basis) and Kramers and Pauli [15] discussed briefly the energy splitting of

* For brevity, components observed (in transverse observation) with their electric vectors perpendicular to the field will be indicated by $\perp$, and those with their electric vectors parallel to the field by $\parallel$.

molecular terms in a magnetic field. The latter investigation not having been seen by Kemble was repeated independently by him [4], [16] in considerably amplified form. Kemble also immediately inaugurated experimental tests of his predictions in the $^1\Sigma \rightarrow {}^1\Pi$ Ångström bands of CO. In the same year Hund [17] deduced expressions for two of his cases of molecular coupling, case (*a*) and case (*b*), while Van Vleck [18] obtained the corresponding results for a rigid molecule on the basis of the quantum mechanics and applied the results to H_2.

§ 2. BEHAVIOUR OF MOLECULES IN A MAGNETIC FIELD

In order to appreciate the significance of the above theoretical work and its later extensions it is necessary to bear in mind that atoms and molecules exhibit distinct effects in the magnetic field for fundamentally similar reasons. In each case we have a mechanical gyroscope (actually not completely rigid in either case) which possesses a certain *magnetic moment*. On account of this magnetic moment an external magnetic field exerts a torque which, because of the gyroscopic forces involved, produces primarily (if the field is not large enough to produce distortion) a steady precession about the direction of the applied field with the well-known Larmor frequency, ω_H, where

$$\omega_H = \frac{\bar{\mu}}{\hbar\sqrt{J(J+1)}} . H \text{ sec.}^{-1} \qquad \ldots.(2)$$

Here $\bar{\mu}$ is the time average of the *magnetic moment* of the atomic or molecular system and J is the integral (or half-integral) rotational quantum number. The vector J determines the total *mechanical* angular momentum of the system in units of $\hbar/2\pi$ according to the relation: (absolute value of vector J) $= \sqrt{J(J+1)}$.* Here we must note that ω_H, as distinguished from the *energy*, is independent of the orientation of the system in the field. Now the contrast between atomic and molecular systems lies in the characteristic differences in the possible values of J and $\bar{\mu}$ and in the effects of molecular rotation on the latter. Thus both in atoms and in non-rotating molecules the range of μ-values is the same. In an atom the mechanical momentum and its measure, $\sqrt{J(J+1)}$, in units of $\hbar/2\pi$ are determined by the mechanical rotation and spin of the active electrons; hence J cannot increase without our having in general a definite increase in μ. In molecules, however, the mechanical rotation and spin of the analogous electrons is so related to μ only for the non-rotating system. As soon as the nuclei (neglecting any small magnetic effect of the nuclei themselves) begin to rotate, the J-vector becomes a vector sum of the angular momentum of the electronic system and that of the relatively heavy atomic nuclei. Hence the quantum number J may increase to huge values (100 or over), with the consequent lack of magnetic

* In what follows there will be repeated need to distinguish between the *vector* J, representing *vector* angular momentum, and the purely scalar quantum number J. To avoid the use of bold-faced type (or some symbol such as J^*) for this vector (and others) it will, where confusion might arise, simply be called the vector J or the J-vector. Where it is necessary to designate the cosine, say of the angle between the J-vector and some other such as the spin-vector, S, for example, this will be written simply as $\cos(J, S)$ so that no confusion need arise.

sensitivity of rigid molecules in high rotational states. This decrease in magnetic sensitivity is further hastened with increasing J by the averaging effect on μ of the molecular rotation itself. (See figure 1 (i).)

As for the average magnetic energy ΔE, we have

$$\Delta E = \bar{\mu} \cos(\mu, H) H,$$

where the bar indicates a time average. Since in general the magnetic-moment vector, μ, of the molecule is carried around by rotation at a rate which is in general high compared with the Larmor precession, we shall, for low fields at

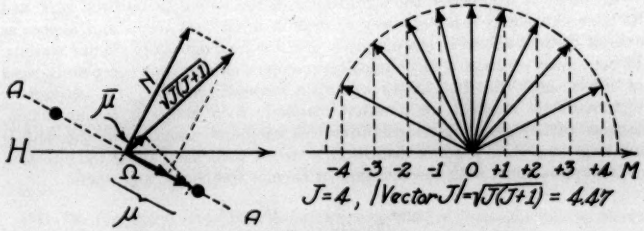


Figure 1. Behaviour of a molecule in the magnetic field.

(i) The total *orbital* and *spin* angular momentum of the excited electrons due to precession around the nuclear axis A–A produces a steady component (averaged over electronic frequencies) along this axis of Ω, where Ω takes on definite quantized values. The angular momentum, N, of nuclear rotation (which is not quantized) then compounds vectorially with Ω to give the total molecular angular momentum represented by the vector J around which they both precess. This vector J has then a magnitude which is given by $\sqrt{J(J+1)}$. Furthermore, associated with Ω is a magnetic moment μ ($=m$ Bohr magnetons) which, being carried around by the rotating molecule, gives an average value (over a rotational cycle) of $\bar{\mu}$ ($=\bar{m}$ Bohr magnetons). It is on the presence of a non-vanishing $\bar{\mu}$ that molecular sensitivity to a magnetic field depends.

(ii) The vector J in an external magnetic field (here assumed horizontal) precesses about the field H at angles such that its projection on H is integral (or half-integral).

least, have an average magnetic moment, $\bar{\mu}$, along the direction of the J-vector. But by the principle of space quantization only those precessional angles are allowed for which the projection of the J-vector on H is integral (or half-integral). Hence, calling the projection, M, the *magnetic quantum* number, we have $(2J+1)$ precessional states, where M varies from $+J$ to $-J$ (figure 1 (ii)). Consequently,

$$\Delta E = \bar{\mu} \cos(J, H) H = \bar{\mu} \frac{M}{\sqrt{J(J+1)}} H. \qquad \ldots.(3)$$

We may now write equation (3) in a form which is at once more convenient and better suited for comparison with the atomic expression by (1) dividing by hc to obtain ΔF, the frequency shift in cm^{-1} of the spectroscopic term value, and (2) expressing $\bar{\mu}$ as a certain average number $\bar{m}$ of Bohr magnetons, where one Bohr magneton is $eh/4\pi m_e c$ ($=0{\cdot}918 \times 10^{-20}$ erg gauss^{-1}).

This gives us, using equation (1),

$$\Delta F = \frac{\Delta E}{hc} = M \frac{\overline{m}e}{4\pi m_e c\sqrt{J(J+1)}} . H = \frac{M\overline{m}}{\sqrt{J(J+1)}} \Delta\nu_n, \qquad \ldots . (4)$$

or

$$\Delta F = M \left[\frac{\text{Average magnetic moment in Bohr magnetons along } J\text{-vector}}{\text{Total mechanical angular momentum}} \right] . \Delta\nu_n. \qquad \ldots . (5)$$

Hence, the *number* of magnetic sub-levels given by a single no-field term depends on J, while their *spacing* is fixed by the ratio in the brackets. In the case of atoms the ratio in the brackets is simply Landé's splitting factor, g, which, being a simple ratio of magnetic to mechanical properties of electrons alone, is of the general order of magnitude of unity (when not zero).

In the case of molecules, as was pointed out earlier, μ has the same general range of values as for atoms. As molecular rotation sets in, however, the vectors J and μ become more and more nearly perpendicular to one another, and μ, being carried around with the rotating nuclei, tends to produce self-cancellation over a cycle. Thus $\bar{\mu}$ decreases rapidly with J. Hence, with the numerator decreasing and the denominator increasing, the g-factors for molecules become very small indeed for large J-values. For this reason magnetic fields produce in molecular energy levels a multiplicity of sub-levels and a fineness of scale which has no direct counterpart in atomic spectra.

To obtain equation (5) in a form suitable for calculation it is of course necessary to carry out correctly the averaging process to find the true average value of the vector μ along the vector J. And it is here that the diversity of results in molecular states originates. If we designate the values of the electronic and spin angular momentum vectors by L and S respectively, then $\bar{\mu}$ depends, not only on their sizes and relative orientations, but also on their relative coupling energies with one another and with the rest of the molecule, and on how these couplings are influenced by rotation and the external magnetic field itself.

§ 3. HUND'S LIMITING MOLECULAR COUPLING TYPES

For our purposes there are three general and two special coupling cases, known as Hund's case (a), case (b) and case (d) on the one hand, and case (b') and case (d') on the other (see figure 2).* In the first three cases we have an electronic orbital angular momentum vector L and an electron spin-vector S, while the last two result as special cases when the S-vector vanishes. As a study of figure 2 illustrates, three of these cases arise when there is vigorous interaction between the axial electric field of the molecule and the L-vector of the rotating electrons. This, in general, causes L to precess more or less rapidly around the nuclear axis in such a way that its projection, Λ, takes on the definite values 0, 1, 2, etc., characteristic of the molecular electronic states designated as Σ, Π, Δ, etc. This precession of the L-vector sets up a definite magnetic moment along the nuclear axis which is the origin of the μ-value for the simplest case, case (b'). The vectors Λ and N in this case

* See also Jevons' *Report* (2), chap. VII, for a more detailed account of these cases.

add to give vector $K(\equiv J)$, the total momentum vector for the system, around which they precess. Here N is not quantized, but takes on such values as, with Λ and J integral, will allow the J-vector to have the size $\sqrt{J(J+1)}$. If we now add the spin-vector S to our picture and the interaction with Λ is *large*,

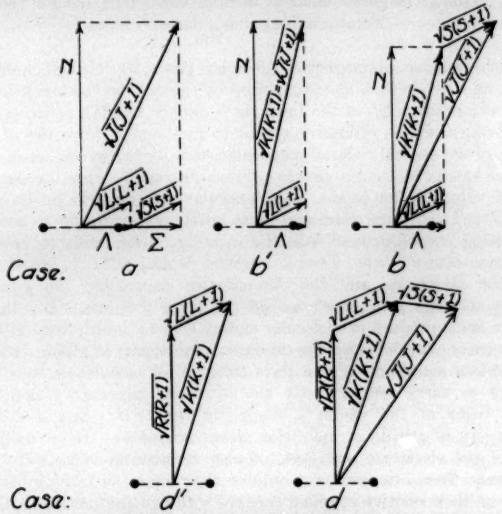


Figure 2. Hund's limiting coupling cases.

Cases (*a*) and (*b*) result when the coupling of the orbital electronic momentum vector, L, with the nuclear axis is *very large*; cases (*d'*) and (*d*) result when this coupling is *negligible* and the magnetic moment set up by nuclear rotation produces coupling of L with R—the nuclear angular momentum vector. If, then, Λ, the projection of the L-vector on the axis, by its magnetic field produces *strong* coupling with the spin-vector S, we have case (*a*); if *weak* coupling case (*b*). Case (*b'*) results from case (*a*) or case (*b*) when spin vanishes, as does case (*d'*) from case (*d*).

so that S is also quantized definitely with respect to the nuclear axis (its projection being Σ), we have the rigid molecule characteristic of case (*a*). Here the vectors Λ and Σ add to give Ω, which in turn compounds with N to give the total momentum vector J. When $\Sigma=0$ we have, therefore, singlet states (case (*b'*) necessarily), when $\Sigma=\frac{1}{2}$ we have doublet states, when $\Sigma=1$ we have triplets, and so on.* If the interaction between S and the axis is small

* It is unfortunate that the same symbol Σ is used to designate (1) the electronic states in the sequences Σ, Π, Δ, etc. for which $\Lambda=0$, and (2) the projections of the spin-vector along the axis. The only justification is the similar use of S in atomic spectra and the strain which the variety of molecular constants puts on *any* system of notation.

(in the limit zero), then, since the rotating nuclei themselves set up a magnetic field along N, this tends to uncouple S from the axis. In this case Λ is added to N to give the quantized vector K around which they precess. S in turn assumes quantized projections along K, $(2S+1)$ in number, which determine the multiplicity of the level (figure 2, case (b)). A special case of some interest arises when $\Lambda=0$, but the spin-vector does not vanish. We then have $^2\Sigma$, $^3\Sigma$, etc. states with the spin-vector quantized along K. In the case of a $^2\Sigma$ state this gives a spin doubling expressible roughly as $\gamma(K+\frac{1}{2})$, where γ is a constant, usually quite small. As a result, this coupling is almost completely broken down at moderate field strengths, giving a Paschen-Back effect, as in atoms. Most molecules, of course, are neither pure case (a) nor (b) but begin near case (a) and gradually approach more or less closely to case (b), depending upon how completely increased rotation is able to break down the spin coupling with the axis.

If now the coupling between the axis and the L-vector in turn is small (in the limit zero) we have rotation gradually uncoupling L from the axis and (in the limit) case-(d') coupling. L is now quantized along the axis of rotation, N, around which it precesses, giving the $(2L+1)$ energy values associated with a so-called "electron complex". Thus, as L-uncoupling proceeds, Λ ceases to have any meaning, and a projection of L perpendicular to the axis of figure, L_{perp}, develops which ultimately tends toward integral values. Actually, all electronic states having a common set of excited electrons, and hence the same L-vector, in the limit of case (d') form a "complex". Thus a $p\sigma\Sigma$, a $p\pi\Pi_e$ and a $p\pi\Pi_d$ state (these last being the close members of a "Λ-doublet") arise from a single p-electron and give in case (d') a *p-complex*, where L_{perp} becomes $+1$, 0 and -1, and K takes on the values $(R+1)$, R and $(R-1)$* respectively. Similarly, in the regular states arising from a single d-electron we have the $d\sigma$ electron tending toward $L_{\text{perp}}=+2$, the $d\pi_e$- and $d\pi_d$-electrons toward $+1$ and 0, and the $d\delta_e$ and $d\delta_d$-electrons toward -1 and -2 respectively †. Thus we see that $\mu\to 0$ for the limiting case (d')-coupling of both a $p\pi\Pi_e$ and a $d\pi\Pi_d$ electronic state.

If we now add electron spin to case-(d') coupling, the spin-vector is coupled to the K-vector by the average value along it of the magnetic moment of L_{perp}. Hence the vectors K and S compound to give the final resultant J-vector, and for each K there are $(2S+1)$ J-values just as in case (b).

§ 4. MAGNETIC SPLITTING OF ENERGY LEVELS IN HUND'S COUPLING CASES

We must now, in terms of the above rather sketchy account, consider the forms that equation (5) (or equation (4)) assumes in these five cases, remembering that, just as with atoms, one Bohr magneton of magnetic moment is to be associated with each *single unit* of *orbital* electronic or each *half-unit* of *spin* electronic angular momentum.

* In this case the momentum of the rotating nuclei along the direction of N is quantized separately (it was not in the first three cases). It is therefore best to give a new designation, the vector R, to this quantized momentum of nuclear rotation.

† See Mulliken [1] Part I (footnote 62).

Case (*a*). Here $m=\Lambda+2\Sigma=\Omega+\Sigma$, and since $\cos(J, \Omega)=\Omega/\sqrt{J(J+1)}$, $\overline{m}=(\Omega+\Sigma)\Omega/\sqrt{J(J+1)}$. Hence equation (5) gives *

$$\underset{\text{case (a)}}{\Delta F}=M\frac{(\Omega+\Sigma)\Omega}{J(J+1)}\Delta\nu_n, \qquad \ldots.(6)$$

and for case (*b'*), since $\Sigma=0$, $\Omega=\Lambda$,

$$\underset{\text{case (b')}}{\Delta F}=M\frac{\Lambda^2}{J(J+1)}.\Delta\nu_n, \qquad \ldots.(7)$$

the result, in different notation, obtained by Van Vleck [18].

Case (*b*). Here we have a marked contrast with the above rigid-molecule behaviour. In fact there are two limiting possibilities, the first obtaining with low fields and relatively strong coupling between the K and S vectors, and the second a *Paschen-Back* effect (quite analogous to the atomic one) when the coupling is so small as to be readily broken down by the field. This Paschen-Back effect arises for all case-(*b*) states of low K-values and for all other states for which the spin-doubling, trebling, etc., is small compared with $\Delta\nu_n$. In this case the spin-vector and the K-vector for the rest of the molecule are *independently* quantized along the field and we have two quantum numbers M_s and M_k, where $M_s=+S, +(S-1)\ldots-S$, and $M_k=+K, +(K-1)\ldots-K$. Hence to this approximation we can write the energies as the sum of two terms of the type of equation (5), giving

$$\underset{\substack{\text{case (b)}\\ \text{Paschen-Back}}}{\Delta F}=\left[\frac{M_k\Lambda^2}{K(K+1)}+2M_s\right]\Delta\nu_n. \qquad \ldots.(7)$$

For low fields where the above breakdown has not occurred we have obviously

$$\overline{m}=\Lambda\cos(\Lambda, K)\cos(K, J)+2\sqrt{S(S+1)}.\cos(S, J),$$

which on substitution in equation (4) and working out of the angles gives (cf. Hund [17]):

$$\underset{\text{case (b)}}{\Delta F}=\frac{M}{J(J+1)}\left[\frac{\Lambda^2}{2K(K+1)}\{J(J+1)+K(K+1)-S(S+1)\}\right.$$
$$\left.+\{J(J+1)-K(K+1)+S(S+1)\}\right]\Delta\nu_n. \qquad \ldots.(8)$$

Case (*d'*). In this case (see figure 2) we have $\overline{m}$ simply as the projection of L on the vector K, which is $\sqrt{L(L+1)}.\cos(K, L)$. Thus, using the law of cosines, and remembering that the magnitudes of the vectors K and R are $\sqrt{K(K+1)}$ and $\sqrt{R(R+1)}$ respectively, we have [19]

$$\Delta F=M\left[\frac{K(K+1)+L(L+1)-R(R+1)}{2K(K+1)}\right]\Delta\nu_n \qquad \ldots.(9)$$

Case (*d*). Since most of the cases of L uncoupling in multiplet states occur where the spin multiplicity is very small, we have complete Paschen-Back effects at all usable fields, and need, therefore, only add (as in equation (7)) $2M_s\Delta\nu_n$ to equation (9) to describe the results adequately.

* After modification in the light of the quantum mechanics this is essentially the result of Hund [17].

§ 5. COMPARISON OF THEORY WITH EXPERIMENT

We may now consider briefly the predictions from the above equations and their comparison with experiment, realizing that they are only *limiting* results and must be modified by quantum-mechanical considerations in all intermediate cases. Since our equations give the magnetic splitting of *single* states and an actual spectrum line results as a transition between *two* states, we must compute Zeeman patterns in general from an appropriate combination of the magnetic sub-levels of both levels. In this, single primes will be used for *initial* and double primes for *final* states, and we must remember that the three main branches of a band result when the changes, ΔJ, in J are $+1$ (P-branch), 0 (Q-branch) and -1 (R-branch).* Further, $\Delta M = \pm 1$ and $\Delta M = 0$ correspond respectively to components polarized $\perp$ and $\parallel$ to the field.

(1) *Zeeman effect in $\Sigma \rightarrow \Sigma$ bands*

Since $m=0$ for a $^1\Sigma$ state we have no magnetic sensitivity, and all $^1\Sigma \rightarrow {}^1\Sigma$ bands should, therefore, be insensitive to the field †, a result verified definitely in the case of the $\lambda 3330$ band of AlH by Watson and Perkins [21]. For $^2\Sigma$ and $^3\Sigma$ states with small spin doubling and trebling we expect uncoupling of the spin-vector from the axis of rotation, with almost complete Paschen-Back effects at moderate fields. Thus, for $^2\Sigma$ states we have from equation (7) with $\Lambda=0$ *two* (and in the case of $^3\Sigma$ *three*) fairly sharp groups of sub-levels of separation $2\Delta\nu_n$ (see figure 3, (iv)) corresponding to $M_s = \pm\frac{1}{2}$ (or $M_s = \pm 1, 0$ in the case of triplets). Now although these levels each contain $(2K+1)$ closely packed sub-levels, we may have only such transitions as are permitted by the Paschen-Back selection rule $\Delta M_s = 0$, i.e. we must have $M_s' = \frac{1}{2} \rightarrow M_s'' = \frac{1}{2}$ and $M_s' = -\frac{1}{2} \rightarrow M_s'' = -\frac{1}{2}$. Hence each line of a band is undisplaced and shows no Zeeman effect whatever. If, now, the spin multiplicity is appreciable, the $(2K+1)$ sub-levels in each of the above groups are somewhat spread out, the exact behaviour requiring precise calculations on the basis of quantum mechanics. These have been carried out by Hill, who obtained the expression (first published by Almy [22]):

$$\Delta F = [\{\nu^2 + 4M\nu\Delta\nu_n\}/\{(K+\tfrac{1}{2}) + 4\Delta\nu_n^2\}]^{\frac{1}{2}}/\mathbf{hc}, \qquad \ldots.(10)$$

where ΔF is the displacement in the field from the doublet mid-point and ν is the doublet separation. This is in harmony with the above prediction when $\nu \ll \Delta\nu_n$, and in the other extreme of $\nu \gg \Delta\nu_n$ it predicts blocks of levels of width $2\Delta\nu_n$ about each no-field spin level. The most commonly reported behaviour in intermediate cases is, however, a second-order effect. A gradual contraction of the lines of a spin-doublet is observed, with the individual lines becoming broader and eventually forming a single diffuse band of width about equal to the original doublet separation. This behaviour in the narrow doublets of N_2^+ and CN has been studied by Deslandres and Burson [11],

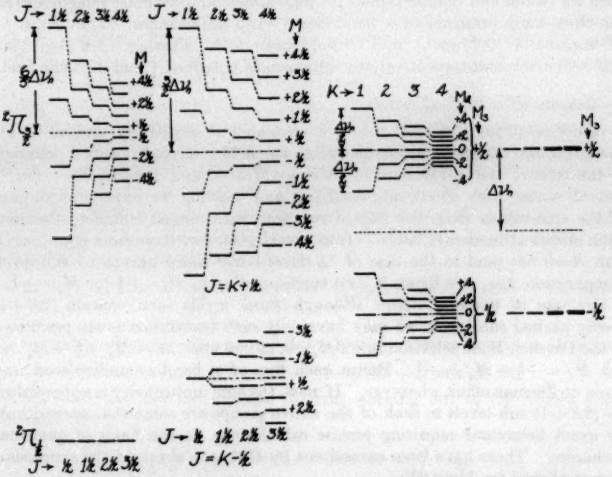
* For details of transitions in more complicated bands see, for example, Jevons' *Report* [2], p. 129 *et seq.*

† This is strictly true only when these states arise from S-electrons or from $p\sigma$ or $d\sigma$ electrons which suffer no L-uncoupling as rotation increases. See, in this connexion, Watson and Bender [20].

by Fortrat [23] and in more detail by Bachem [24], who expresses his extensive results in the form

$$\nu\delta\nu=\alpha H^2\left(1-\frac{\alpha}{\nu^2}H^2+\dots\right), \qquad \dots.(11)$$

where ν=no-field doublet width, $\delta\nu$=contraction produced by the field, and α is a constant. While this result cannot be read directly from equation (10), the author has shown * that after some approximations a result in general harmony with equation (11) is obtained.



(i) $^3\Pi$ case (*a*). (ii) $^3\Pi$ case (*b*). (iii) $^3\Pi$ Paschen-Back. (iv) $^3\Pi$.

Figure 3. (i) and (ii) (after Jevons [22]): Low-field Zeeman levels for $^3\Pi$ case-(*a*) and case-(*b*) states as given by equations (6) and (8). (The horizontal dotted lines represent no-field positions.) (iii) and (iv): Paschen-Back effects in $^3\Pi$ and $^3\Sigma$ states as given by equation (7). (Λ=1 and 0 respectively.)

Calculations based on the quantum mechanics have also been made by Schmid, Budó and Zemplen [25] for $^3\Sigma$ states with *large* spin trebling. The results, while not capable of simple expression except in certain ranges, are in good agreement with their cleverly made absorption measurements on the λ7600 $^1\Sigma\leftarrow{}^3\Sigma$ band of O_2. This behaviour of $^3\Sigma$ states seems also to be confirmed by Croze's [11] observations on the red sub-levels of $^3\Pi\rightarrow{}^3\Sigma$ first positive bands of N_2, where triplet patterns of average width $3{\cdot}99\Delta\nu_n$ are found. This is just what we should expect, since the upper state, $^3\Pi_0$, is magnetically

insensitive (see below), and a complete Paschen-Back effect is expected in the $^3\Sigma$ state at even moderate fields. Similar results are reported by Schmid [26] for a third positive CO band (λ2975) involving a $^3\Sigma \rightarrow {}^3\Pi_0$ transition.

(2) *Zeeman effect in* $^1\Sigma \rightarrow {}^3\Pi$ *and* $^1\Pi \rightarrow {}^1\Sigma$ *bands*

Owing to the insensitivity of the strict case-(b') $^1\Sigma$ states the band types most suited to a rigid test of theoretical predictions are $^1\Sigma \rightarrow {}^1\Pi$ and $^1\Pi \rightarrow {}^1\Sigma$, since any observed effects can be ascribed to the $^1\Pi$ state alone. If, then, bands with a strict case-(b') $^1\Pi$ state (i.e. one with no L-uncoupling) can be used, the applicability of equation (7) is assured. The field of selection is further greatly reduced by the desirability of having (i) a light molecule to

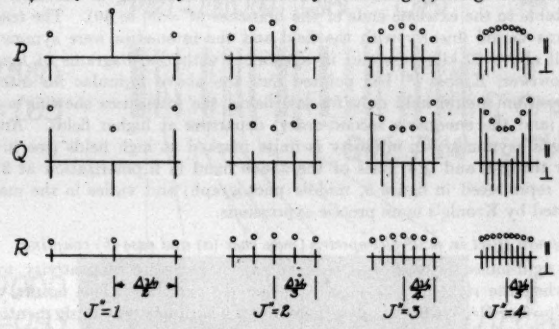


Figure 4. Weak-field Zeeman patterns for early lines of the P, Q and R branches of a $^1\Sigma \rightarrow {}^1\Pi$ band. (After Kemble, Mulliken and the author [28].)

Here $\Delta\nu_n$ is the "normal" half-width of an atomic triplet, and intensities of the $\parallel$ components of each pattern are plotted upward and those of the $\perp$ components downward from the horizontal. The intensities to be expected when no polarizing device is used are indicated by the heights of the small circles above the horizontal lines. These diagrams apply also to $^1\Pi \rightarrow {}^1\Sigma$ bands where the Q lines behave as for $^1\Sigma \rightarrow {}^1\Pi$ bands, while the top row gives the patterns for $R(0)$ $R(1)$ $R(2)$ and $R(3)$ lines respectively and the bottom row those for $P(2)$ $P(3)$ $P(4)$ and $P(5)$ lines respectively.

insure open-line structure, (ii) bands in spectral regions susceptible to work in air with rapid emulsions, and (iii) bands capable of intense excitation with freedom from impurities. These various restrictions have, as a result, limited work to three sets of bands, the $^1\Pi \rightarrow {}^1\Sigma$ band of He_2 and the λ4240 band of AlH, which were investigated qualitatively by Curtis and Jevons [27] and Watson and Perkins [21] respectively, and the $^1\Sigma \rightarrow {}^1\Pi$ Ångström bands of CO, examined in considerable detail in both polarizations by Kemble, Mulliken and the author [28]. In the last case five bands of the CO Ångström system were studied over a series of field strengths up to 36 kilogauss and the theoretical predictions verified *qualitatively* to the smallest details and *quantitatively* in all essential features. The patterns to be expected from equation (7) are

shown diagrammatically in figure 4, where *intensities* and *positions* of components in ∥ and ⊥ polarizations (observation at right angles to the field) are shown in solid lines *above* and *below* the horizontal respectively.* The circles represent by their distance above the horizontal the corresponding intensities when no polarizing medium is present. The intensities are computed from the expressions †, of Kronig [29], Hönl [30] and Dennison [31]. Actually, it was possible to resolve 13 of the possible 18 patterns predicted for J'' (the end quantum number)=1 and 2, and to trace the marked doublet-like patterns of the Q lines in ∥ polarization to $J''=16$ and of the P lines in ⊥ polarization to $J''=10$. The gross overall sizes of the unresolved patterns were also in harmony with expectations, the broadening in the field being detectable to the extreme ends of the branches ($J''=30$ to 40). The resolved patterns varied linearly with the field and the intensities were symmetrical to well above 20 kilogauss and in agreement with the diagrams of figure 4. As, however, Kronig [32] has pointed out, the above formulae for intensity and position should hold only for low fields, the intensities showing a first-order (and the energies a second-order) departure at higher fields. And, in fact, this asymmetry in intensity is quite marked at high fields (see in particular the $Q(3)$ and $Q(4)$ lines of the λ5610 band in ∥ polarization at 34 880 gauss, reproduced in figure 5, middle photograph) and varies in the manner predicted by Kronig's more precise expressions.

(3) *Zeeman effect in multiplet spectra (pure case-(a) and case-(b) coupling)*

In rigid-molecule case-(a) states we expect complete insensitivity to the field when the right-hand side of equation (6) vanishes. This occurs when $|\Omega+\Sigma|=0$ or $\Omega=0$, and includes, therefore, the $^1\Sigma$ states previously mentioned ($\Lambda=\Sigma=0$ and the magnetic moment $m=0$), $^2\Pi_{\frac{1}{2}}$ and $^3\Delta_1$ states (where the magnetic moments of Λ and Σ are equal and opposite and cancel) and, finally, $^3\Pi_0$ states. In this last case $\Omega=0$ while $|\Omega+\Sigma|=1$. This means that the *mechanical* momenta (Ω) of Λ and Σ cancel, thus causing the J-vector to be along N (figure 1 (i)). Hence the magnetic moment μ, of one Bohr magneton, is perpendicular to J and cancels completely over a rotational cycle.

The general behaviour of the magnetic sub-levels as calculated from equations (6), (7) and (8) is best illustrated diagrammatically and is given for the lowest rotational states of $^2\Pi$ levels in figure (2) (i, ii, iii). Here the decrease of overall width with increase of components is quite striking. We see now the reason for the $2\Delta\nu_n$ separation of the double heads reported in the field by Dufour [6] in the case of the $^2\Pi\rightarrow{}^2\Sigma$ bands of the alkaline-earth halides, and later (CaF, SrF, BaF, with better experimental arrangements) by von Mathes [33]. In all these bands we have an upper state which is near case (a) and a lower state with very small spin doubling, so that each $^2\Sigma$ rotational level gives the usual sharp doublet of separation $2\Delta\nu_n$ even for high J-values. (See figure 3 (iv).)

* Here, just as in the atomic case, transitions with $\Delta M=\pm1$ are observed in ⊥ polarization and only those with $\Delta M=0$ in ∥ polarization (observation being, of course, at right-angles to H).

† These are given in the present notation in convenient form in the author's *Review* [5] as equations (8), p. 95.

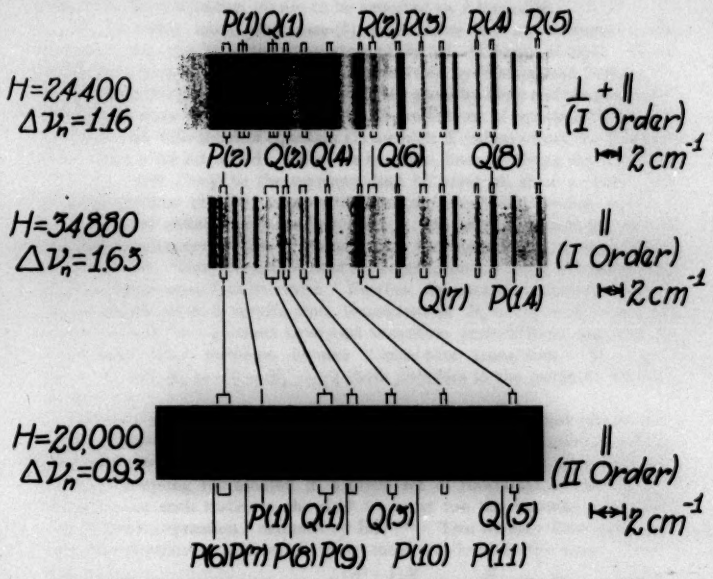


Figure 5. Reproduction of portions of plates of the λ5610 band of CO [28] in the field.

These represent enlargements of about 12 diameters made from plates representing exposures of from 24 to 48 hours' duration in the first and second orders of a 21-foot Rowland concave grating with 20 000 lines per inch. To be particularly noted are the clearly resolved triplets of the $P(1)$ and $Q(1)$ lines in the upper photograph (both polarizations superposed); the single undisplaced component constituting the pattern of the $P(1)$ line and the wider doublet that of the $Q(1)$ line in ∥ polarization; and finally the pronounced doublet character of early Q lines in ∥ polarization. At medium fields (bottom photograph at 20 000 gauss) the Q doublets are symmetrical in intensity, while at higher fields the *low*-frequency member (on the left in the photographs) is distinctly more intense (middle photograph $Q(3)$ and $Q(4)$). This is in harmony with predictions of Kronig's perturbation calculations [32].

[*To face p.* 166

Further, the heads of these bands are formed by lines involving very high J-values, and for these the upper states are nearly insensitive to the field. Hence results the doubling of the heads by the observed amount. In other $^2\Pi$ states the molecule is near case (*a*) *only*, for the lowest rotational states and departures from equation (6) are to be expected as J increases.

On the other hand pure case-(*b*) states have few representatives, the nearest to this type being the upper state of the $^2\Pi \rightarrow {}^2\Sigma$ bands of MgH. These bands have been studied in the field qualitatively by Watson and Perkins [21], and quantitatively at fields from 5000 to 28 400 gauss by Almy and the author [34]. The results show near agreement with the predictions of equation (8) for low K-values and with those of equation (7) for high K-values where the Paschen-Back effect is far advanced. Thus, for example, lines involving the transition from $^2\Pi_{\frac{1}{2}}$ (with $J'=\frac{1}{2}$) to the corresponding $^2\Pi$ level all show a fairly symmetrical pattern of eight components (as expected) and overall width of $4{\cdot}4\Delta\nu_n$ (where equation (8) predicts $5{\cdot}0\Delta\nu_n$). At large K-values, in contrast, we have symmetrical *triplets* of overall width very nearly $4\Delta\nu_n$, as equation (7) would predict. Here the members of the triplets are not sharp, but are roughly of the original spin-doublet width. Further, the central component is much the strongest since it results from transitions of $M_s'=+\frac{1}{2} \rightarrow M_s''=+\frac{1}{2}$ and $M_s'=-\frac{1}{2} \rightarrow M_s''=-\frac{1}{2}$ (these have *high* transition probabilities), and the two very weak outer members involve "cross-over transitions" $M_s'=+\frac{1}{2} \rightarrow M_s''=-\frac{1}{2}$ and $M_s'=-\frac{1}{2} \rightarrow M_s''=+\frac{1}{2}$ (both *forbidden* in the complete Paschen-Back case).

Actually the vast majority of doublet-type states represent intermediate cases. Thus they may lie somewhere near case (*a*) for low rotation and pass gradually toward case (*b*) as rotation increases, with, therefore, all intermediate types of coupling represented in a given set of rotational levels. For the discussion of such states in the field we must use the rigorous (and rather complicated) expression * deduced by Hill [35]. This we may write, measuring from the rotationless state of the given molecular level, in the form

$$F(J)_H = B[J(J+1)-(\Lambda\pm\tfrac{1}{2})^2+\tfrac{1}{4}]+B\frac{(\Lambda^2+\frac{1}{2})M}{J(J+1)}\cdot\Delta\nu_n \pm \frac{B}{2}\sqrt{(\theta+\delta^2)}, \quad \ldots.(12)$$

where the first term on the right does not contain the field and there are several such in the radical of the last term (it being, of course, necessary that this equation should reduce to Hill and Van Vleck's general formula for doublet molecular levels when H vanishes).

Here the + sign refers to the F_1 levels with $J=(K+\frac{1}{2})$ and the − sign to the F_2 levels with $J=(K-\frac{1}{2})$. Also B is the well-known band-spectrum constant $\hbar/8\pi^2cI$ cm.$^{-1}$, and θ and δ are abbreviations. Explicitly

$$\theta = 4(J+\tfrac{1}{2})^2(1-\beta)^2+\Lambda^2(Y+\beta).(Y-4+5\beta),$$

where $\beta=\dfrac{M}{BJ(J+1)}\cdot\Delta\nu_n$ and Y is a coupling parameter, where $Y=\infty$ corresponds to pure case-(*a*) and $Y=0$ or $+4$ to pure case-(*b*) states.†

* This neglects spin-doubling in $^2\Sigma$ states and "Λ-doubling" in $^2\Sigma$ states.

† See Mulliken and Christy [36], Table II. Actually $Y=A/B$, where A is the spin-doubling constant and gives the separation between the $^2\Pi_{\frac{1}{2}}$ (Λ and Σ parallel) and the $^2\Pi_{\frac{3}{2}}$ (Λ and Σ antiparallel) states when no rotation is present.

The expression for δ is quite cumbersome though, fortunately, it may be neglected for large Y-values.* Actually the third term in equation (12), subtracting such terms as have nothing to do with the field, reduces for *large* Y-values to $\frac{3}{2}\frac{M}{J(J+1)}.\Delta\nu_a$, which is precisely the value of the second term! Hence in the case-(a) limit this predicts the identical results of equation (6),

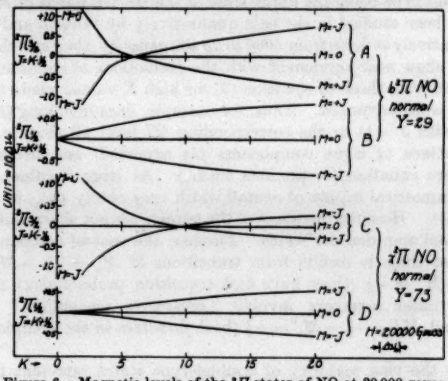


Figure 6. Magnetic levels of the $^2\Pi$ states of NO at 20 000 gauss.

The extreme sub-levels $M=\pm J$ and the fictitious $M=0$ levels are here drawn from Hill's equation (12), where, since δ is practically negligible, the curves are essentially symmetrical. The peculiar crossing-over of the levels in A and C and the gradually increasing sensitivity of the $^2\Pi_{\frac{1}{2}}$ levels as shown in B and D is characteristic and in marked contrast to the predictions of either equations (6) or (8). Thus in A (or C) we begin with our molecule near case (a) with Λ and S parallel. As rotation increases, however, we pass towards case-(b) coupling, and S tends more and more towards anti-parallelism to K. At some intermediate point the average magnetic moments of Λ and S along K cancel exactly with resulting magnetic insensitivity. This point of cross-over of course occurs at higher K values the greater the value of Y. Likewise in B (or D) rotation gradually uncouples S from the axis towards parallelism to K and hence Λ and the initially anti-parallel S no longer cancel magnetically.

namely: $3M\Delta\nu_a/J(J+1)$ for a $^2\Pi_{\frac{1}{2}}$ state and zero effect for a $^2\Pi_{\frac{3}{2}}$ state. As Y decreases this radical differs more and more from the above simple contribution, causing in general the $^2\Pi_{\frac{1}{2}}$ state to become *less* sensitive and the $^2\Pi_{\frac{3}{2}}$ state *more* sensitive as K increases. The physical reason for the behaviour is quite clear. When Y is large we are, in the low rotational $^2\Pi_{\frac{1}{2}}$ state, near case (a), with the S-vector tending towards parallelism with Λ, with which it is coupled. As rotation increases, however, the S-vector experiences increased coupling with the K-vector, tending more and more towards anti-parallelism with it. Hence at some K-value the average magnetic moment

* For its complete form in our notation see Crawford (5), p. 104, equations (32), (33), (34) and (35). The θ used above stands for $(\epsilon_1-\epsilon_2)^2$ in that *Review*.

of S along K is *equal* and *opposite* to the projection of the moment due to Λ. Cancellation then results, $\overline{m}=0$, and the state is magnetically insensitive for this value of K. Above this K-value $\bar{\mu}$ is anti-parallel to K and we have a "crossing over" of the magnetic energy levels. This is illustrated clearly in figure 6 for the $b^3\Pi$ and $x^3\Pi$ states of NO, where the crossing-over occurs at larger K-values as Y is increased. The increasing sensitiveness of the $^3\Pi_1$ state is likewise caused by the gradual breaking away of the S-vector (which begins anti-parallel to Λ) and its finally becoming *parallel* to K.

For $^3\Pi_1$ states $|Y|>50$, if we take $J=\frac{1}{2}$ and $H=30\,000$ gauss, the predictions of equation (12) differ by less than 4 per cent from those of equation (6), while for larger J-values and smaller Y-values the contributions of the radical, and in particular δ, become dominant. In such cases we must use the complete expression where the calculations become quite tedious.

The detailed results to be expected from Hill's theory have been compared with experiment by a great number of observers in a wide variety of coupling cases between (*a*) and (*b*) and over a large range of the magnetic field. This includes work on the $^2\Pi\rightarrow{}^2\Sigma$ band of MgH at $\lambda5211$ ($Y'=5$) by Watson and Perkins [21] and Almy and the author [34]; the $^2\Sigma\rightarrow{}^2\Pi$ band of OH at $\lambda3064$ by Deslandres and d'Azambuja [37] and Fortrat [38] and, in much greater detail, Almy [22]; the $^3\Pi\rightarrow{}^3\Sigma$ bands of ZnH ($Y'=46{\cdot}7$), and HgH ($Y'=570$) by Hulthén [12] and Watson and Perkins [21], and those of ZnH and on the similar bands of CdH ($Y'=167$) by Watson [39]; the $^2\Sigma\rightarrow{}^2\Pi$ (γ) bands of NO ($Y''=73$) by Fortrat [40], Pogany and Schmid [41] (discussed by Almy [42] in the light of Hill's theory); the $b\,^2\Pi\rightarrow x\,^2\Pi$ (β) bands of NO by Schmid [43]; the $^2\Sigma\rightarrow{}^2\Pi$ band of CH at $\lambda3900$ studied by Fortrat [44] (discussed by Almy [42]); and finally the $^2\Pi\rightarrow{}^2\Sigma$ comet-tail bands of CO^+ by Schmid and Gerö [45]. Here in general the theory seems to agree quite satisfactorily with the experimental results over the whole range of values of K, Y and H studied, provided, of course, that (1) spin-doubling is not too great at large K-values for equation (12) to be able to describe it accurately when $H=0$, (2) exposures are sufficiently protracted to give the fainter (but significant) portions of patterns, and (3) the state concerned, beginning as intermediate case (*a*) and (*b*), does not experience L-uncoupling with increased rotation. This last type of behaviour has actually been found by Watson and Bender [20] in their examination of the $^2\Pi\rightarrow{}^2\Sigma$ band of CaH at $\lambda7000$. Here $Y'=18{\cdot}5$ and Hill's theory fits the earlier lines quite well, after which there is progressive departure as L-uncoupling proceeds.

(4) *Zeeman effect in case-(d') states (L-uncoupling)*

The effects to be expected in case-(*d*) states with very *small* spin multiplicity are essentially those predicted for (*d'*) singlet states and need not be considered separately. Now the significant feature about equation (9) is that, except when $K=R$, the coefficient of $M\Delta\nu_n$ remains very large for all K-values, with the result that much larger Zeeman effects may now be expected in the higher lines of all branches (with the exception of those deriving from $p_m\Pi_e$ levels since here $L_{perp}\rightarrow0$ as L-uncoupling proceeds).

Thus, for example, equation (9) for a $d\sigma\Sigma$ state (with $L_{perp}=+2$ and $K=R+2$) gives $\Delta F=2M\Delta\nu_n/K$ and hence an overall width of the magnetic sub-levels of $4\Delta\nu_n$. Similarly, when $L_{perp}=-2$ an overall width of $2\Delta\nu_n$ is predicted, these overall widths being the same for all K-values.

In the case of H_2, where L-uncoupling is very marked in the singlet states, the fragmentary results indicate indeed that the effects are greatest in lines involving $d\sigma\,{}^1\Sigma$ levels. Thus in the $3d\sigma\,{}^1\Sigma \rightarrow 2p\sigma\,{}^1\Sigma$ lines we have uncoupling in both states with L_{perp} tending toward $+2$ in the upper and $+1$ in the lower

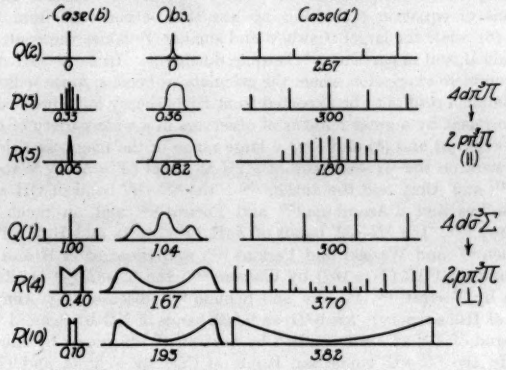


Figure 7. Observed patterns of He_2 lines compared with pure case (b') and case (d') theoretical predictions. (After Mills [49].)

Here are represented a few of the more striking patterns obtained in the $4d\pi\,{}^3\Pi \rightarrow 2p\pi\,{}^3\Pi$ and $4d\sigma\,{}^3\Sigma \rightarrow 2p\pi\,{}^3\Pi$ bands of He_2, where the final state in each case is known to be normal case (b'), while both initial states involve marked L-uncoupling (and hence approach towards case (d')) as rotation increases. Hence the patterns observed agree more closely with the pure case-(b') patterns on the left for low rotation and resemble those on the right more and more as rotation increases. The predictions of Knipp's more exact expressions are in reasonable agreement in most of these intermediate patterns—where differences occur they are probably due to failure to take into account the intensity distribution in unresolved patterns.

state. Hence with " blocks " of levels of width $4\Delta\nu_n$ for the *upper* and $2\Delta\nu_n$ for the *lower* state we expect patterns of overall width $2\Delta\nu_n$ for lines of all K. This seems to be in agreement with Dufour's observations [46], where the sensitive lines do in fact give the rough width of $2\Delta\nu_n$. On the other hand, the observations of Dufour [47] and Croze [8] on the Fulcher bands (which are $p\pi\,{}^3\Pi \rightarrow s\sigma\,{}^3\Sigma$) show no appreciable effect. Hence we conclude that the upper state can show no marked uncoupling since the lower is magnetically insensitive. (We may neglect here the similar Paschen-Back splitting of both levels since it gives *lines* in the no-field position only.)

In the case of He_2 where various degrees of L-uncoupling are present, bands of "orthohelium" with a d-complex as upper state and the $2p\pi\,{}^3\Pi$ as a lower have been studied by Curtis and Jevons [27], Harvey [48] and Mills [49]. This work has verified in a striking fashion the above prediction of uniformity of pattern width for all high values of K in the case of $nd\sigma\,{}^3\Sigma \rightarrow 2p\pi\,{}^3\Pi$ and $nd\pi\,{}^3\Pi_e \rightarrow 2p\pi\,{}^3\Pi$ bands. For the $nd\pi\,{}^3\Pi_f \rightarrow 2p\pi\,{}^3\Pi$ bands, however, the effect is large at the origin and falls off rapidly with K, which is just what we should expect since, as case-(d') coupling is reached, L_{perp} tends towards *zero*. Figure 7 illustrates in a striking fashion some of the cases studied by Mills in which as K increases we pass from excellent agreement with case-(b') predictions to closer approximation to those of case (d'). This situation, intermediate between cases (b') and (d'), has been discussed theoretically by Knipp [50], whose expressions and curves illustrate in p- and d-electron complexes the transitions from one type of coupling to the other. The overall widths determined from these curves agree in general with Mills' quantitative data, such discrepancies as there are being probably due to intensity distributions and impossibility of obtaining well resolved patterns in all cases. In particular they show the "crossing-over" in the 3- and 4-$d\delta\Delta$ states of He_2, occurring when the component of the magnetic moment in case (b') changes its direction, as case (d') is approached, to *anti*-parallelism with K. This produces complete insensitivity to the field at some intermediate value of K for the same reason as in the transition from case (a) to case (b) (see figure 6).

In the case of doublet states with *large* spin doubling, Watson and Cunningham [51] have found marked indication of L-uncoupling for high K-values in the $4p\pi\,{}^2\Pi \rightarrow {}^2\Sigma$ and $4p\lambda\sigma\,{}^2\Sigma \rightarrow {}^2\Sigma$ bands of CaH where approximate correction for the known values of L_{perp} gives better agreement with observations than Hill's theory alone.

In conclusion it may be reiterated that magnetic effects and their study have played a lesser rôle in the elucidation of molecular than atomic spectra. They have none the less been of great value. In general they have brought out in a striking fashion some of the characteristic similarities and differences in molecular and atomic systems, and their detailed verification of theoretical predictions has furnished a gratifying demonstration of the fundamental soundness of the physical ideas underlying the whole of present-day band-spectrum theory.

REFERENCES

(1) Mulliken, R. S. *Rev. Mod. Phys.* **2**, 60 (1930); **3**, 89 (1931); **4**, 1 (1932).
(2) Jevons, W. "Report on Band Spectra of Diatomic Molecules" (*London: Phys. Soc.*, 1932).
(3) Zeeman, P. *Verslag. K. Akad. Wet., Amst.*, **5**, 248 (1896); *Astrophys. J.* **5**, 32 (1897); *Phil. Mag.* **43**, 226 (1897).
Becquerel, J. and Deslandres, H. *C.R. Acad. Sci., Paris*, **127**, 18 (1898).
Michelson, A. A. *Nature, Lond.*, **59**, 440 (1899).
Purvis, J. E. *Proc. Camb. Phil. Soc.* **13**, 354 (1906).
(4) Kemble, E. C. "Molecular Spectra in Gases." *Bull. Nat. Res. Coun., Wash.*, **57**, ch. vii (1927).

(5) Crawford, F. H. *Rev. Mod. Phys.* **6**, 90 (1934).
(6) Dufour, A. *Ann. Phys., Paris*, **9**, 409 (1908); *C.R. Acad. Sci., Paris*, **140**, 118, 229, 810 (1908); *Le Radium*, **5**, 86, 291 (1908); *J. Phys.* **8**, 237 (1909).
(7) Becquerel, J. *C.R. Acad. Sci., Paris*, **146**, 683 (1908).
(8) Croze, M. F. *Ann. Phys., Paris*, **1**, 63, 127 (1914).
(9) Fortrat, R. *C.R. Acad. Sci., Paris*, **156**, 1459 (1913); *Ann. Phys., Paris*, **3**, 282 (1915).
(10) Deslandres, H. and d'Azambuja, L. *C.R. Acad. Sci., Paris*, **157**, 671 (1913).
(11) Deslandres, H. and Burson, V. *C.R. Acad. Sci., Paris*, **157**, 1105 (1913); **158**, 1851 (1914).
(12) Hulthén, E. Dissertation, Lund (1923).
(13) Fortrat, R. *Ann. Phys., Paris*, **3**, 282 (1915).
(14) Lenz, W. *Verh. dtsch. phys. Ges.* **21**, 632 (1919).
(15) Kramers, H. A. and Pauli, W. *Z. Phys.* **13**, 351 (1923).
(16) Kemble, E. C. *Phys. Rev.* **27**, 799 (1926).
(17) Hund, F. *Z. Phys.* **36**, 657 (1926).
(18) Van Vleck, J. H. *Phys. Rev.* **28**, 1006 (1926). See also Kronig, R. de L., *Z. Phys.* **46**, 814 (1927).
(19) Mills, J. S. *Phys. Rev.* **38**, 1148 (1931).
(20) Watson, W. W. and Bender, W. *Phys. Rev.* **35**, 1513 (1930).
(21) Watson, W. W. and Perkins, B. *Phys. Rev.* **30**, 592 (1927).
(22) Almy, G. H. *Phys. Rev.* **35**, 1495 (1930).
(23) Fortrat, R. *C.R. Acad. Sci., Paris*, **158**, 334 (1914).
(24) Bachem, A. *Z. Phys.* **3**, 372 (1920).
(25) Schmid, R., Budó, A., and Zemplén, J. *Z. Phys.* **103**, 250 (1936).
Schmid, *Phys. Rev.* **49**, 271 (1936).
(26) Schmid, R. *Phys. Rev.* **39**, 539 (1932).
(27) Curtis, W. E. and Jevons, W. *Nature, Lond.*, **116**, 746 (1925); *Proc. Roy. Soc.* A, **120**, 110 (1928). See also Mulliken, R. S. and Monk, G. S., *Phys. Rev.* **34**, 1530 (1929).
(28) Kemble, E. C., Mulliken, R. S. and Crawford, F. H. *Phys. Rev.* **30**, 438 (1927).
Crawford, F. H. *Phys. Rev.* **33**, 341 (1929).
(29) Kronig, R. de L. *Z. Phys.* **31**, 885 (1925).
(30) Hönl, H. *Z. Phys.* **31**, 340 (1925).
Hönl, H. and London, F. *Z. Phys.* **33**, 803 (1925).
(31) Dennison, D. M. *Phys. Rev.* **28**, 318 (1926).
(32) Kronig, R. de L. *Phys. Rev.* **31**, 195 (1928).
(33) Mathes, E. von. *Z. Phys.* **68**, 493 (1931).
(34) Almy, G. H. and Crawford, F. H. *Phys. Rev.* **34**, 1517 (1929).
(35) Hill, E. L. *Phys. Rev.* **34**, 1507 (1929).
(36) Mulliken, R. S. and Christy, A. *Phys. Rev.* **38**, 87 (1931).
(37) Deslandres, H. and d'Azambuza, L. *C.R. Acad. Sci., Paris*, **157**, 814 (1913); **158**, 153 (1914).
(38) Fortrat, R. *J. Phys. Radium*, **5**, 20 (1924).
(39) Watson, W. W. *Phys. Rev.* **36**, 1134 (1930).
(40) Fortrat, R. *C.R. Acad. Sci., Paris*, **157**, 991 (1913); and *Ann. Phys., Paris*, **3**, 345 (1915).
(41) Pogany, B. and Schmid, R. *Z. Phys.* **49**, 162 (1928).
(42) Almy, G. F. Thesis, Harvard University (not published *in extenso*).
(43) Schmid, R. *Z. Phys.* **59**, 42 (1929).
(44) Fortrat, R. *Ann. Phys., Paris*, **19**, 81 (1923).
(45) Schmid, R. and Gerö, L. *Z. Phys.* **86**, 314 (1933).
(46) Dufour, A. *C.R. Acad. Sci., Paris*, **146**, 634 (1908); and *Phys. Z.* **10**, 124 (1909).

(47) Dufour, A. *Ann. Chim. Phys., Paris* (8), **9**, 361 (1907); *J. Phys.* **8**, 258 (1909).
(48) Harvey, A *Proc. Roy. Soc.* A, **126**, 583 (1929).
(49) Mills, J. S. *Phys. Rev.* **37**, 1005 (1931).
(50) Knipp, J. K. *Phys. Rev.* **47**, 672 (1935).
(51) Watson, W. W. *Phys. Rev.* **39**, 278 (1932).
Cunningham, W. P. *Phys. Rev.* **41**, 389 (1932).
Cunningham, W. P. and Watson, W. W. *Phys. Rev.* **44**, 815 (1933).

THE LUMINESCENCE OF SOLIDS

By J. T. RANDALL, D.Sc., and M. H. F. WILKINS, B.A.
University of Birmingham

§ 1. INTRODUCTION

During the last few years interest in this subject has quickened a great deal [1], [2], and progress has come about largely as a result of more accurate experimental work, coupled with theoretical developments on electronic energy states in solids [3]. Commercial interest in the application of luminescent materials to lighting problems has also been a very considerable stimulus [4]. The present article is written from the general point of view, as it is not possible in the space available to discuss each aspect in detail.

The term " luminescence " is used in a general sense to indicate the emission of radiation by solids in the visible and near-visible region as a result of the absorption of ultra-violet, x rays, cathode rays or α particles. One of the

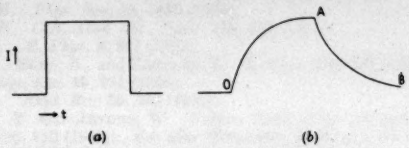


Fig. 1. Intensity of radiation as a function of time.

most striking properties of luminescent solids is associated with the time delay of emission. Absorption of radiation by atoms of a gas may lead to emission in as brief an interval after absorption as 10^{-8} sec. In a solid, however, the emission process may spread over a period of hours. Figure 1 (*a*) represents the time-variation of intensity of a beam of ultra-violet radiation, and figure 1 (*b*) represents diagrammatically the same thing for a luminescent solid. It is to be noted that the luminescence not only takes an appreciable time to decay but also an appreciable time to reach equilibrium. The emission of radiation by the solid during excitation is usually referred to as " fluorescence " ; if the duration of luminescence after removal of the exciting source is of a greater order than 10^{-6} sec. it is generally referred to as " phosphorescence ".

Two chief classes of luminescent solids are now recognized, and the more restricted of these is concerned with pure substances such as platinocyanides, the uranyl salts, and certain compounds of elements with incomplete $3d$ or $4f$ shells, such as manganese and the rare earths. Where purity of materials is at issue no single line of evidence is particularly convincing, but the following considerations serve to reduce uncertainties to a minimum :—

(i) If the fluorescent salt is soluble, the solution may also be fluorescent : this test applies to uranyl salts, platinocyanides and pure rare-earth salts.

(ii) The decay law is exponential in type over the whole range, and the time constant small.

(iii) Pure fluorescent solids are not photo-conducting, at least at low temperatures.

So far as is known at present, fluorescence in pure solids is to be associated either with coordination groups, with molecules, or with ions containing incomplete groups such as those mentioned above [4], [5]. Transitions take place

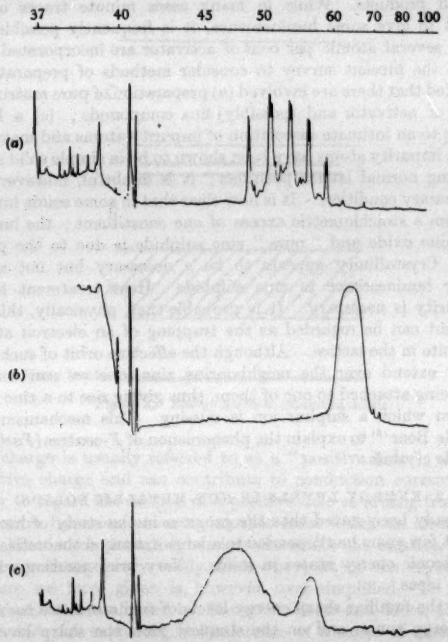


Figure 2. Spectra of various luminescent solids. The lines showing on the left are in each case due to the exciting mercury radiation.

(*a*) Uranyl nitrate, 90° K. (*b*) Manganese chloride, 90° K. (*c*) Zinc sulphide (copper impurity), 300° K.

within the group or the ion, and in such cases comparatively sharp emission spectra may be obtained (see figure 2 (*a*) and (*b*)). In such compounds the emitting centre is behaving to some extent as a free atom or molecule and is little influenced by the presence of near neighbours. The electron distribution

in coordination groups would in general be complicated, and it is quite possible for diffuse spectra to be obtained; it is thought that the emission of tungstates may be of this type.

Most luminescent solids, however, owe their properties to the presence of an impurity, as was first pointed out by Boisbaudron [7]. Thus, copper, silver and manganese are well known "activators" of phosphors of the zinc and zinc-cadmium sulphide series; bismuth is commonly used with commercial alkaline-earth products. While in many cases minute traces of impurity are sufficient to give some luminescence, it is frequently possible to obtain results when several atomic per cent of activator are incorporated. There is not space in the present survey to consider methods of preparation, but it should be noted that there are involved (*a*) preparation of pure matrix material; (*b*) addition of activator and (possibly) flux compounds; (*c*) a heat treatment, leading to an intimate association of impurity atoms and matrix lattice; in some cases impurity atoms have been shown to be in simple solid solution [8], thus occupying normal lattice positions; it is doubtful, however, if this is always a necessary condition. It is now clear that in some solids luminescence may arise from a stoichiometric excess of one constituent; the luminescence of "pure" zinc oxide and "pure" zinc sulphide is due to the presence of excess zinc. Crystallinity appears to be a necessary but not a sufficient condition for luminescence in zinc sulphide. Heat treatment to produce the self-impurity is necessary. It is probable that, physically, this excess of one constituent can be regarded as the trapping of an electron at a vacant negative-ion site in the lattice. Although the effective orbit of such a trapped electron may extend over the neighbouring zinc ions we can conveniently regard it as being attached to one of them, thus giving rise to a zinc atom near to a site from which a sulphur ion is missing. This mechanism has been proposed by de Boer [9] to explain the phenomenon of *F*-centres (*Farbenzentren*) in alkali-halide crystals.

§ 2. ENERGY LEVELS IN NON-METALLIC SOLIDS

It has already been stated that the progress in the study of luminescence during the last few years has depended to a large extent on theoretical developments of electronic energy states in solids. Very brief mention will now be made of these ideas.

Instead of the familiar sharp energy levels of free atoms we have in solids a series of energy zones, and on the simplest view the sharp levels of the free atoms become broadened into bands by aggregation into the solid state. In a metal the uppermost band containing electrons is only partly full, but in a non-conductor the corresponding band A is full and is separated from the next possible region of energy levels by a forbidden zone B. In order to reach band C, electrons in A must absorb a quantum of energy at least equal to the minimum energy difference between band A and band C. When an electron is raised to band C it is said to be in the conduction band of the crystal and is separated from the parent atom and thus free; this process is in this sense equivalent to the ionization of a free atom. A crystal in which electrons may

be raised to the conduction band by the absorption of a quantum of radiation is a photo-conductor. It should be noted that the X-ray levels in solids remain comparatively sharp; they are largely uninfluenced by the field of surrounding ions. The work of Skinner [11] on the transition of electrons between the full band A and X-ray levels is of particular interest. The transition of an electron either to an X-ray level (emission) or to the conduction band (absorption) leaves behind a "hole" in the full band A. This absence of

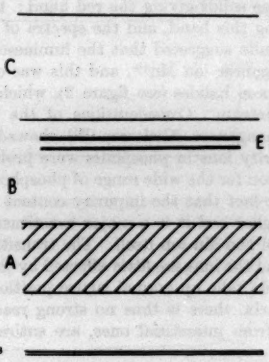


Figure 3. Energy levels in insulating solids.

negative charge is usually referred to as a "positive hole", since it behaves as a positive charge and can contribute to conduction currents. Physically it is easier to regard the motion of a positive hole as arising from the jumping of an electron from one atom to another from which an electron is already missing. Since such potential barriers are small, the location of the missing electron may change very easily from one part of the crystal to another. The picture we have given is, however, over-simplified. It was shown by Peierls [5], and later confirmed by Slater [10], Seitz and Mott, that there were strong grounds for supposing the existence of a series of excitation states E below the conduction band. An electron raised from a full band A to an excitation state is a bound electron and is still in the field of the parent atom. A series of bound stationary states exists for the positive hole and electron leading up to some series limit. A bound electron and positive hole may move through the crystal, and the system has been called an "exciton" by Frenkel [12]. The excitation levels may belong to the pure crystal (e.g. alkali-halide measurements by Schneider and O'Bryan [13]) or to impurity centres.

§ 3. THE FUNCTION OF THE IMPURITY

It is a remarkable fact that out of all the possible impurities in the periodic table the element manganese should be the one to give luminescence in the greatest number of cases. The function of this impurity has been investigated in detail by Randall [14] for a large number of solids. Kröger [15] has since studied the case of zinc sulphide using large percentages of manganese. Although solids containing manganese impurity are known to luminesce in the yellow, green and red, the clue to the general function of manganese was obtained from those solids giving the red band; there are in fact large numbers of solids giving this band, and the spectra of a few such are shown in figure 4. These results suggested that the luminescence was due to transitions within the manganese ion Mn^{++}, and this was confirmed by measurements on pure manganese halides (see figure 2), which showed substantially the same emission spectrum. Considerations of the absorption spectra of pure compounds of manganese (Gielessen [16]) showed that the transitions within the Mn^{++} impurity ions in phosphors were probably of the forbidden type ${}^2G \rightarrow {}^4F$. The reason for the wide range of phosphors in which manganese is to be found, and the fact that the impurity content can range over values so wide that at the higher end it is a major constituent of the phosphor, is due to the properties of the Mn ion itself. The transitions take place within the shielded 3*d* shell, and the ions are little affected by the nature of the lattice. While manganese ions do take up normal lattice positions, giving rise to solid solutions with the matrix, there is thus no strong reason for assuming that such ions, as distinct from interstitial ones, are entirely responsible for the observed effects.

Considered from the point of view of the Bloch and Wilson treatments of energy levels in solids, it is clear that the impurity levels are in many instances equivalent to the excitation levels E already mentioned. Normally it would be expected that increasing the impurity content beyond, say, 1 in 10^2 or 1 in 10^3 would have marked effects on the luminescence; that is to say, on the relative dispositions of energy bands and excitation levels. This is observed in many cases. It is only possible, apparently, to use large amounts of impurity when the atom concerned has a well-shielded shell. Even so, the optimum amount is not usually greater than 1 per cent. (e.g. 0·5 per cent for Mn in Zn_2SiO_4 and 2 per cent Mn in ZnS). The impurity levels may extend into the conduction band, in which case photo-conduction should be possible at all temperatures.

Self-impurities have already been mentioned. So far little experimental evidence of value exists regarding these, but it is clear that a small number of atoms of one of the constituents must be different from the rest of its kind.

§ 4. PHOSPHORESCENCE

The time delays of emission already referred to have long been known in a qualitative way. Owing to the design of the apparatus and uncertainties with regard to impurity content, it is doubtful if any of the early measurements

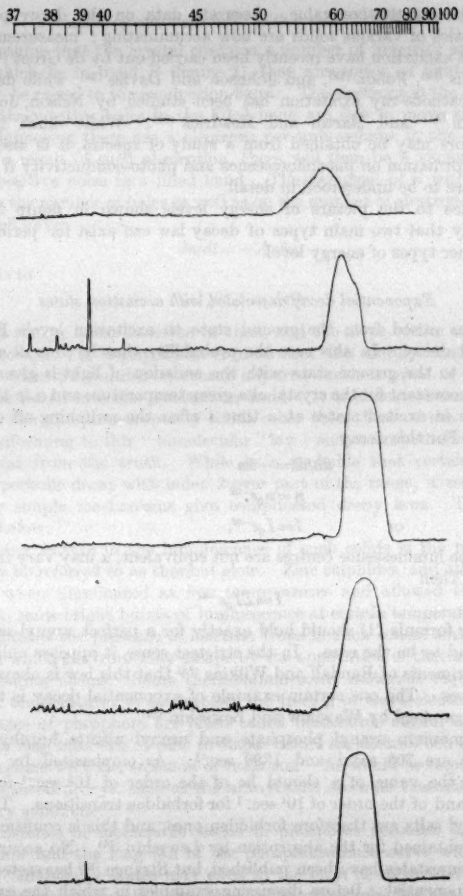


Figure 4. Fluorescence spectra of solids with manganese impurity, 90° K.
(Reproduced by permission of the Council of the Royal Society.)
The curves refer in order to: Cadmium sulphate + 0·1 % Mn, magnesium oxide + 1 % Mn, zinc sulphide + 0·01 % Mn, zinc pyrophosphate + 0·1 % Mn, magnesium silicate + 0·1 % Mn, cadmium borate + 0·1 % Mn.

have much quantitative value. Accurate data on the decay periods and characteristics of various solids are now accumulating; measurements under ultra-violet excitation have recently been carried out by de Groot [17], Randall and Wilkins [18], Fonda [19], and Johnson and Davies [20], while decay subsequent to cathode-ray excitation has been studied by Nelson, Johnson and Nottingham [21] and Martin and Headrick [22]. While much information on phosphors may be obtained from a study of spectra, it is also necessary to have information on phosphorescence and photo-conductivity if any of the processes are to be understood in detail.

Reference to the picture of energy levels shown in figure 3 indicates immediately that two main types of decay law can exist for perfect crystals with no other types of energy level.

Exponential decay associated with excitation states

Electrons raised from the ground state to excitation levels E result in exponential decay. In this case the probability that in time dt an electron will return to the ground state with the emission of light is given by $\alpha n\,dt$, where α is a constant for the crystal at a given temperature and n is the number of electrons in excited states at a time t after the switching-off of exciting radiation. For this decay

$$dn/dt = -\alpha n$$

and

$$n = n_0 e^{-\alpha t}$$

or

$$I = I_0 e^{-\alpha t}. \qquad \ldots.(1)$$

When the luminescence centres are not equivalent, α may vary from centre to centre. Then

$$I = \Sigma I_0 e^{-\alpha t}.$$

The simpler formula (1) should hold exactly for a perfect uranyl-salt crystal. This is found to be the case. In the strictest sense it appears unlikely from recent experiments of Randall and Wilkins [23] that this law is obeyed, except in a few cases. The one certain example of exponential decay is the uranyl salts, also examined by Wawilow and Lewschin [24].

For ammonium uranyl phosphate and uranyl nitrate hexahydrate the values of α are 703 sec^{-1} and 1390 sec^{-1}. As emphasized by Mott and Gurney [25], the value of α should be of the order of 10^8 sec^{-1} for allowed transitions and of the order of 10^4 sec^{-1} for forbidden transitions. Transitions in the uranyl salts are therefore forbidden ones, and this is confirmed by the low values obtained for the absorption by Lewschin [26]. No accurate decay curves for tungstates have been published, but Strange [27] has stated that the decay is exponential. Before discussing examples in which the exponential law is only approximately obeyed [18], [20], it will be necessary to consider what happens when electrons and positive holes are separated after the absorption process.

Phosphorescence associated with conduction levels

If we assume that the crystal contains a number of impurity atoms, some levels of which lie in band B (figure 3), then absorption of light will cause electrons to be raised to the conduction band. The electrons in the conduction band and the positive holes in the filled band are free to move through the crystal. Supposing there are n electrons per unit volume in the conduction band; as a result of such absorption, there will then be n vacant impurity levels or positive holes in a filled band. The variation of n with time now depends on the number of holes as well as on the number of electrons in band C, and we shall have

$$dn/dt = -An^2,$$

which leads to

$$I = I_0/(1+n_0At)^2. \qquad \ldots.(2)$$

$1/\sqrt{I}$ is thus a linear function of time, and the slope of the curve is proportional to $\sqrt{I}$. It is important to note that this simple bimolecular law holds only when the number of vacant impurity-levels is equal to the number of electrons raised into the conduction band. It is almost common practice to refer to the phosphorescence of solids other than those of the exponential type as conforming to this "bimolecular" law; such a simplified statement is, however, far from the truth. While it is probable that certain sulphides show a hyperbolic decay with index 2 over part of the range, it seems certain that quite simple mechanisms give complicated decay laws. This will be amplified below.

A striking feature in the luminescence of such solids is the phenomenon which may be referred to as thermal glow. Zinc sulphides, and alkaline-earth sulphides when illuminated at low temperatures and allowed to warm up in the dark, show bright bursts of luminescence at certain temperatures. This clearly indicates the presence of electron traps in such solids, small potential holes from which electrons may escape by the application of thermal vibration of neighbouring atoms. Some thermoluminescent materials require moderately high temperatures to accomplish the freeing of these electrons. These energy states of phosphors are often called metastable states, and make it possible for long intervals of time to elapse before an electron can return to an impurity centre with the emission of radiation. As would be expected, solids with exponential decays, such as the uranyl salts, show no thermal glow, or at most a very small one.

The existence of metastable states in phosphors accounts both for the thermal glow and the long tail of the phosphorescence curve which is often observed. Metastable states may exist in the form of vacant lattice sites or Schottky defects. In such a case there exist more holes for the electrons to return to than there are free electrons.

Using the previous notation, with b as the number of additional vacant holes,

$$dn/dt = -An(n+b),$$

which gives (Wawilow [28])

$$n=b/(Be^{bAt}-1), \quad \text{where} \quad B=1+b/n_0.$$

Hence with t large the decay is exponential, but for t small if $n_0>b$ the decay is hyperbolic. Also in [17] de Groot gives a discussion of the decay for a photo-conductor with traps in the conduction band.

In deriving equation (2) for luminescent decay in a photo-conductor it is assumed that the electron falls directly from the conduction band to the ground state. If, however, the electron first passes into an excitation level before it drops to the ground state, the life-time of the excitation state will affect the decay curve. For a short life-time the curve will be hyperbolic except that for high intensities the initial quick decay will be exponential, with a time constant characteristic of the excitation level. Mott and Gurney [25] suggest that the initial very fast exponential decay for tungstates and sulphides reported by Strange [27] is to be accounted for in this way. The same explanation might also hold for the almost exponential initial decay of zinc orthosilicate, although it is also possible that some of the electrons raised to excitation states are not carried into the conduction band [22]. An alternative explanation of the willemite decay (Fonda [19]) is that all the electrons pass into the conduction band and return to the ground state without delay in an excitation level. The initial exponential decay is due to a large excess of positive holes over electrons owing to the greater number of electrons being trapped in the conduction band. The alkali halides activated by thallium have been discussed by Seitz [29], who considers that metastable states within thallium molecules account for the long exponential decay.

From the above discussions it is clear that simple deduction of the nature of the phosphorescence mechanism cannot always be made from the phosphorescence decay curve. It is therefore important that other lines of approach be also used, such as the investigation of photo-conduction currents. The measurement of quantum efficiency for photo-conduction and luminescence is basic to the problem, and the variation of decay curves and photo-conduction with temperature is another important method.

§ 5. ABSORPTION PROBLEMS AND INTENSITY EFFECTS

The absorption peaks on the long-wave-length side of the fundamental absorption band in pure alkali halides have already been mentioned [5], [13]. Peaks have also been observed in impure solids and are characteristic of the impurity. Thus Koch [30] showed that the absorption coefficient for one of the exciting bands in KCl+TlCl is approximately proportional to the thallium concentration. Kröger [15] has shown the existence of new absorption peaks in zinc sulphide and zinc orthosilicate due to the presence of manganese. Whereas the solid solution of zinc sulphide and cadmium sulphide produces a shift of the absorption edge to longer wave-lengths with increasing cadmium content, the addition of manganese produces new bands, which would be expected from the known properties of the manganese ion [14].

Although such features have been observed, it cannot be stated that the measurement of absorption coefficients in phosphors is in a satisfactory state. Practically all measurements have been made on powders or thin films, and are for this reason qualitative. The question of the fundamental absorption edge of zinc sulphide is still in dispute; Seitz [31] considers the edge to be in the far ultra-violet (~2500 A.), while the Dutch workers [15] consider the edge to be at 3380 A. It is likely that the difficulties in this crystal are produced to some extent by the unknown amount of self-impurity present; the near ultra-violet tail may in fact be due to the self-impurity. Satisfactory results are likely to be obtained only by the use of single crystals.

In general the intensity of luminescent light from a solid is proportional to the intensity of the exciting light. Absorption seems to be in the main a single-stage process, whereas in luminescent gases a molecule may often absorb two quanta in rapid succession, and such double excitation produces luminescence proportional to the square of the exciting intensity. In phosphors containing a small percentage of impurity the luminescence intensity tends to saturate for high exciting intensities (de Groot [34]). The probability of a photon being absorbed at an impurity centre is reduced when an appreciable proportion of the centres is at any given time in the excited state. Thus the saturation effect can be accounted for in terms of the absorption coefficient (which decreases with high intensities), the concentration of centres and the mean life-time of an excited centre. In the case of the uranyl salts the number of luminescence centres is large and no saturation can be detected; strictly proportional intensity relations hold. In the case of the zinc sulphides the long-period phosphorescence and thermal glow are readily saturated at low intensities because the number of traps which can hold excited electrons is limited. This gives rise to non-linear intensity relations for low intensities. Consider a sulphide which emits bands A and B. Two kinds of electrons can fall from the conduction band, those which have just emerged from traps and which prefer to get rid of their energy via band A, and those electrons which have not entered traps and which do not show this preference. As the exciting intensity increases, the number of electrons emerging from traps tends to a limit, and as a result the proportion of electrons emitting via band A to those emitting via band B becomes less. The total emission from A and B is proportional to the exciting intensity, hence the intensity of band A is less than proportional, and that of B more than proportional, to this intensity (Kröger and Gisolf [32], Riehl [33]).

§ 6. THE APPLICATIONS OF LUMINESCENT SOLIDS

The chief applications of luminescent solids at the present time are concerned with discharge-tube lighting and devices suitable for mitigating the dangers of black-out conditions. Other important applications are concerned with cathode-ray tubes either for television reception or for laboratory oscillographs [4]. Luminescent solids have also long been used for theatrical purposes. While most of these applications make use of the emission of light during

excitation, war-time conditions give a new importance to luminescent powders with long afterglows. For this reason luminescent solids possessing in high degree the metastable states already discussed are of great value at the present time.

In spite of the early experiments of Becquerel on the excitation of luminescence in solids by means of a low-pressure discharge, no serious proposals were made in this field until 1923, when Risler [35] suggested that such fluorescent solids as those of the zinc sulphide class could be usefully applied to the inside and outside of high-tension positive-column glow discharges. Difficulties in applying the powders in suitable uniform layers to the inside walls of the tubes led to the use of powders on the outside of the tubes, but little success was achieved in this way. It was shown in 1932 that uniform layers of appropriate thickness could be produced on the inside wall of the discharge tube [36]. The method consisted in forcing along the tube a bung moistened with an appropriate binder, the necessary amount of binder being extremely small. Dry powder of appropriate particle size was then allowed to flow over the adhesive surface, thus forming a thin layer. The earlier use of zinc sulphide in this way was discontinued in low-pressure tubes when it was found that such compounds as zinc orthosilicate, zinc beryllium silicate [37] and cadmium chlorophosphate [33] were more efficiently excited by the shorter wave-lengths characteristic of this type of discharge. The luminous efficiency of the low-pressure mercury discharge can be increased from 5 lumens/watt to 55 lumens/watt by the use of zinc orthosilicate activated by manganese, while the zinc beryllium silicates give 30–40 lumens/watt under the same conditions. The colour of the "zinc silicate" tubes is bright green, while those coated with zinc beryllium silicates may range from creamy-white to pink and from yellow-green to orange. Such tubes are largely used for advertising signs and for interior lighting in special circumstances. Some of these powders are also excited by neon discharges [39]. In the low-pressure type of discharge it is fair to say that the chief advantage of employing luminescent coatings is a gain of efficiency. Such powders are also applied to the inside of the outer bulbs of the high-pressure mercury lamps used for street lighting, and here the advantage is chiefly a modification of the colour, the efficiency remaining substantially the same [40].

In cathode-ray tubes for television purposes sulphides with short afterglow are chiefly used. In order to obtain a "white" screen, complementary combinations of zinc and zinc-cadmium sulphides are employed. Characteristics of powders under electron bombardment have recently been studied in detail by Nottingham [41], and Nelson, Johnson and Nottingham [21].

Interesting details of luminescent powders for radiographic applications and zinc sulphides containing radioactive material have been given by Levy and West [42], who also make some reference to phosphors with long decay periods. It is impossible to discuss these numerous applications in detail, and reference should be made both to the accounts already cited which deal specifically with applications, and to the patent literature.

REFERENCES

(1) *Acta phys. Polon.* **5** (1936).
(2) *Trans. Faraday Soc.* **35** (1939).
(3) Wilson. *Proc. Roy. Soc.* A, **133**, 450 (1931).
Fowler. *Proc. Roy. Soc.* A, **140**, 505 (1933).
(4) Randall. *J. Roy. Soc. Arts*, **85**, 354 (1937).
Ryde. *Light and Lighting*, **31**, 137 (1938).
Davies. *Trans. Faraday Soc.* **35**, 171 (1939).
(5) Peierls. *Ann. Phys., Lpz.*, **13**, 905 (1932); *Annales de l'Institut H. Poincaré*, **5**, 177 (1936).
Mott. *Proc. Roy. Soc.* A, **167**, 384 (1938).
Frenkel. *Phys. Z. Sowjet*, **9**, 158 (1936).
(6) Randall. *Trans. Faraday Soc.* **35** (1939).
(7) Boisbaudron. *C.R. Acad. Sci., Paris*, **103**, 468 (1886).
(8) Rooksby, McKeag and Jenkins. *Nature, Lond.*, **143**, 978 (1939).
(9) de Boer. *Rec. Trav. Chim. Pays-Bas*, **56**, 301 (1937).
(10) Slater and Shockley. *Phys. Rev.* **50**, 705 (1936).
(11) Skinner. *Rep. Progr. Phys.* **5**, 257 (1938).
(12) Frenkel. *Phys. Z. Sowjet*, **8**, 185 (1935); **9**, 158 (1936).
(13) Schneider and O'Bryan. *Phys. Rev.* **51**, 293 (1937).
(14) Randall. *Proc. Roy. Soc.* A, **170**, 272 (1939); *Nature, Lond.*, **142**, 113 (1938).
(15) Kröger. *Z. Kristallogr.* A, **100**, 543 (1939); *Physica*, **6**, 369 778, (1939).
(16) Gielessen. *Ann. Phys., Lpz.*, **22**, 537 (1935).
(17) de Groot. *Physica*, **6**, 275 (1939).
(18) Randall and Wilkins. *Nature, Lond.*, **143**, 978 (1939).
(19) Fonda. *J. Appl. Phys.* **10**, 408 (1939).
(20) Johnson and Davies. *J. Opt. Soc. Amer.* **29**, 283 (1939).
(21) Nelson, Johnson and Nottingham. *J. Appl. Phys.* **10**, 335 (1939).
(22) Martin and Headrick. *J. Appl. Phys.* **10**, 116 (1939).
(23) Randall and Wilkins. (In the press.)
(24) Wawilow and Lewschin. *Z. Phys.* **48**, 397 (1928).
(25) Mott and Gurney. *Electronic Processes in Ionic Crystals.* (Oxford Univ. Press, 1940.)
(26) Lewschin. *Acta Physico chimica, U.S.S.R.*, **6**, 661 (1937).
(27) Strange. *Trans. Faraday Soc.* **35**, 95 (1939).
(28) Wawilow. *Phys. Z. Sowjet*, **3**, 369 (1934).
(29) Seitz. *J. Chem. Phys.* **6**, 150 (1938).
(30) Koch. *Z. Phys.* **57**, 638 (1929); **59**, 378 (1930).
(31) Seitz. *Trans. Faraday Soc.* **35**, 99 (1939).
(32) Gisolf and Kröger. *Physica*, **6**, 1101 (1939).
(33) Riehl. *Z. tech. Phys.* **5**, 152 (1939).
(34) de Groot. *Physica*, **6**, 393 (1939).
(35) J. Risler. British Patent 208,723 (1923).
(36) ,, British Patent 421, 932 (1933).
(37) McKeag and Randall. British Patent 480,356 (1936).
(38) ,, ,, British Patent 495,706 (1938).
(39) Jenkins. British Patent 457,486 (1935).
(40) Ryde. British Patent 401,846 (1932); 422,364 (1934).
(41) Nottingham. *J. Appl. Phys.* **8**, 762 (1937).
(42) Levy and West. *Chem. Ind. Rev.* **58**, 457 (1939).

REACTIONS IN SOLIDS

By N. F. MOTT, F.R.S.
H. H. Wills Physical Laboratory, University of Bristol

§ 1. INTRODUCTION

Some explanation may be needed of the inclusion of an article about chemical reactions in a volume on progress in physics. My excuse is that, in order to understand reactions in the solid state, one has to make use of conceptions which are not chemical in the ordinary sense, and they might just as well be called physics as anything else. For instance, the crystal structures of the reacting materials are usually important, and the crystal structures of metals and inorganic compounds have generally been considered to be a branch of physics. Then, especially in reactions in metals, the shapes of the precipitates formed by the reaction may be important, because of the strains which they set up in the material and the resulting influence on its mechanical properties, such as hardness and yield point. Further, in considering the oxidation of aluminium at room temperature, we have to consider why a protective layer some 50 A. thick is formed, after which the reaction stops. The explanation of this turns out to be a typically quantum-mechanical phenomenon.

One purpose of this article is to compare reactions which occur in metals, such as those in which a new phase is precipitated from a supersaturated solution (e.g. $CuAl_2$ from a solid solution of Cu in Al), with those in which an ionic crystal is involved. In the latter class we include the oxidation of metals, the photo-chemical reduction of silver halides and the thermal decomposition of metallic azides. Many of these reactions involving ionic substances were discussed at a recent meeting of the Faraday Society [1]; their theory will be developed in detail in a forthcoming book by R. W. Gurney and the present author [2].

In any chemical reaction involving two or more chemically different solid substances—that is to say, excluding mere changes of crystal structure, diffusion must take place. The first section, therefore, will be devoted to a discussion of the mechanism of diffusion in solids. After that, we treat reactions between two solid phases, their rate of variation with temperature, and the strains set up in the medium. Finally we discuss the conditions under which reactions will start, and thus the question of induction periods and the formation of nuclei.

§ 2. DIFFUSION AND ELECTROLYTIC CONDUCTION IN SOLIDS

We shall begin by discussing self-diffusion in solids. This has been measured in lead [3], copper [4] and for the lead ion in PbI_2 [5] by the method of radioactive indicators. We may distinguish between lattice diffusion proper, which takes

place in a perfect crystal and is independent of its past history, and structure-sensitive diffusion. The coefficient of self-diffusion varies approximately according to the law

$$D = D_0 e^{-Q/RT}. \qquad \ldots.(1)$$

For lattice diffusion proper, at least three mechanisms have been proposed by which diffusion can take place; it is quite possible that all three of them occur in nature in different cases. They are:

(1) *A simple exchange of atoms, as illustrated in figure* 1. The activation energy Q will then be the energy required to bring a pair of atoms half-way round, as illustrated in figure 1 (*b*).

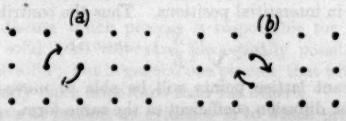


Figure 1. Illustrating a pair of atoms changing places in a solid.
In (*a*) they are just starting, in (*b*) they are half-way round.

(2) *The Frenkel mechanism of interstitial ions and holes.* According to Frenkel [6], in any crystal in thermodynamical equilibrium at a temperature T, a certain proportion of the atoms (or ions) will have left their normal lattice positions and moved into interstitial positions as shown in figure 2 (*a*). In, for instance, a face-centred cubic or in the simple cubic rock-salt structure,

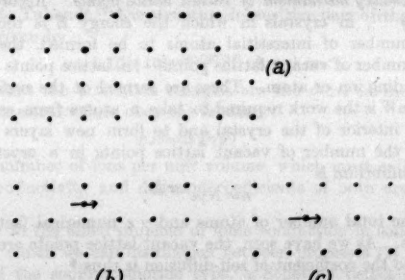


Figure 2. (*a*) illustrates a Frenkel lattice defect, i.e. an interstitial atom and a vacant lattice point; (*b*) the motion of an interstitial ion; (*c*) that of a vacant lattice point.

the interstitial positions will be the centres of the cubes. If E is the energy required to move an atom from its normal position to an interstitial position, it may be shown that the number n of atoms in interstitial positions, and hence the number of vacant lattice points, is given by

$$n = \gamma\sqrt{(NN')}\, e^{-\frac{1}{2}E/kT}. \qquad \ldots.(2)$$

Here N is the total number of atoms, N' the total number of possible interstitial positions, and γ a numerical factor depending on the change in the vibrational frequencies of the atoms adjacent to holes.

The interstitial atoms will be able to move through the crystal. A certain activation energy U will, however, be necessary to bring an atom to the half-way position (figure 2 (b)). The diffusion coefficient of an interstitial atom will then be of the form

$$D_i = D_{i0} e^{-U/kT}.$$

The contribution to the coefficient of self-diffusion due to the interstitial ions will be obtained by multiplying D_i by n/N, the fraction of the total number of ions which are in interstitial positions. Thus the contribution to D is

$$D_{i0}\gamma\sqrt{\left(\frac{N'}{N}\right)}e^{-(\frac{1}{2}E+U)/kT}.$$

Similarly the vacant lattice points will be able to move, and will make a contribution to the diffusion coefficient of the same form,

$$D_0'\gamma\sqrt{\left(\frac{N'}{N}\right)}e^{-(\frac{1}{2}E+U')/kT}, \qquad \ldots.(3)$$

where U' is the energy necessary to move an atom half-way into a hole (figure (2c)). If U and U' are sensibly different, one exponential factor will be very much greater than the other; we must thus identify Q with $(\frac{1}{2}E+U)$ or $(\frac{1}{2}E+U')$, whichever is the smaller.

(3) *The Schottky mechanism of vacant lattice points.* According to Wagner and Schottky [7], in crystals in which the energy E is too great for an appreciable number of interstitial atoms to be formed, there may nevertheless be a number of vacant lattice points—i.e. lattice points unoccupied by the corresponding ion or atom. These are formed at the surface and diffuse inwards. If nW is the work required to take n atoms from separated lattice points in the interior of the crystal and to form new layers on the crystal surface, then the number of vacant lattice points in a crystal in thermodynamical equilibrium is

$$n = N\gamma e^{-W/kT},$$

where N is the total number of atoms and γ a numerical factor of the same type as before. As we have seen, the vacant lattice points are mobile; their contribution to the coefficient of self-diffusion is thus *

$$D = D_{0i}\gamma e^{-(W+U')/kT}.$$

We must emphasize that in actual materials the proportion of vacant lattice points and interstitial atoms is rather small, probably not more than one in 10^5 even just below the melting point. At low temperatures (room temperature in rock salt, liquid air in silver bromide) the mobility becomes vanishingly small and a number of holes or interstitial ions may get frozen-in; their number will thus exceed the equilibrium value.

* For the application of the idea of vacant lattice points to liquids, cf. a number of papers by Lennard-Jones and others [8].

There are two important differences between process (1) on the one hand and processes (2) and (3) on the other. These are :

A. Process (1) cannot lead to any transport of matter. The specimen will be the same shape at the beginning as at the end. This will not necessarily be the case for processes (2) and (3).

B. In ionic crystals the lattice points of the crystal are occupied by ions rather than atoms, and so in processes (2) and (3) we have to do with interstitial ions, and lattice points from which an ion is missing. Thus the diffusion process is accompanied by the motion of electrical charge. Thus processes (2) and (3) lead to ionic conductivity in the solid state, while process (1) does not

We must now discuss which process is responsible for diffusion in any particular type of solid. All three are theoretically possible, and may all take place in any one solid ; but in general one process, that with the smallest Q, will be overwhelmingly more important than the others.

In metals we have at present no evidence as to which process takes place, though (2) seems unlikely for closely packed metals.

For ionic crystals it is probable that (2) or (3) is responsible. For one case, the lead ions in lead iodide, we have quite definite evidence that diffusion is due either to process (2) or to (3), i.e. to a process which leads to a movement of charge. The diffusion constant of the lead ions has been found [5] to vary with the temperature according to the formula

$$D = 10{\cdot}6\, e^{-30{,}000/RT}\ \mathrm{cm^2/sec.},$$

while according to Tubandt [9] the structure-insensitive part of the electrolytic conductivity is given by

$$\sigma = 1{\cdot}15 \times 10^5 e^{-30{,}000/RT}\ \mathrm{ohm^{-1}\ cm^{-1}}$$

It will be seen that the activation energies are the same in the two cases ; moreover the numerical values obey the relation

$$\sigma/D = Ne^2/kT, \qquad \ldots.(4)$$

where N is the number of ions per unit volume, which must exist between σ and D if the conductivity and diffusion coefficients of both are due to the same process.

Turning now to the allied problem of ionic conduction in solids, there are various ways in which we can distinguish between processes (2) and (3). In ionic crystals of the sodium chloride type, a Schottky defect must clearly consist of a *pair* of vacant lattice points, from which ions of opposite sign have been removed. Thus the number of vacant lattice points of each sign must be equal. We may expect their mobilities to be of the same order of magnitude. Thus if process (3) is the important one for conduction, we should expect *both* ions to be mobile—the transport numbers for both constituents would be of the same order of magnitude. On the other hand, the energy necessary to put a metal ion into an interstitial position will in general be less than that required for the larger negative ion ; and so if process (2) is the important one, only the metal ions will be mobile.

Process (3) (Schottky defects) will be favoured in cases where the two ions are of nearly the same size ; process (2) (Frenkel defects) when one ion is small, and when they are polarizable. It is believed that process (3) is mainly responsible for ionic conduction in alkali halides, process (2) in silver halides.

As regards the energies E, W and U, attempts to make theoretical estimates of them have been made by Jost [10], and by Mott and Littleton [11], for alkali halide crystals. We may remark here that W, the energy required to form a vacant lattice point, is equal to W_L, the lattice energy per molecule, in those crystals in which the molecules are bound together by van der Waals forces (e.g. solid rare gases). For ionic crystals the calculations referred to above show that $W < W_L$. For metals no calculations have been made.

Up to the present we have discussed only self-diffusion and ionic conductivity in pure metals and in ionic crystals of stoichiometric composition. We now have to consider the diffusion of one substance, A, into another, B. If the

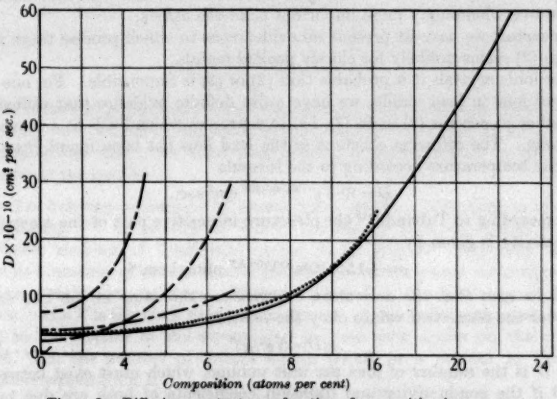


Figure 3. Diffusion constant as a function of composition in copper alloys (Rhines and Mehl) at 800° c.

Al ....; Be – – –; Si – – – –; Sn – – – –; Zn —.

substance A forms with the substance B a substitutional solid solution, atoms of B replacing atoms of A in the crystal lattice, then diffusion may be accounted for by any one of the mechanisms already postulated to account for self-diffusion. This will in general be the case for the diffusion of one metal into another, or one ionic crystal into another.

In the case of metals the diffusion constant D depends in general on the composition. Figure 3 (due to Rhines and Mehl [12]) shows the diffusion constant for a number of metals in copper. The fact that the diffusion constants tend to about the same values for low concentrations is striking ; it suggests that the mechanism of diffusion is (3) above, the diffusion rate being mainly dependent on the number of vacant lattice points present at any temperature.

In cases where the diffusing substance goes into solution interstitially the mechanism of diffusion must necessarily be interstitial also, the diffusing atom moving from one interstitial position to another. The diffusion constant will vary with temperature according to the formula

$$D = D_0 e^{-U/kT},$$

where U is the energy already defined required to activate the jump from one interstitial position to another. Examples of interstitial solution are carbon in iron and hydrogen in palladium. In many cases, also, ionic crystals can dissolve an excess of metal interstitially; this will be discussed in the next section.

We must now say a few words on structure-sensitive diffusion. In metals, grain-boundary diffusion is well known, and always proceeds with a lower activation energy than lattice diffusion. For instance, for thorium in tungsten, Q for grain-boundary diffusion is 94 kcal. and for lattice diffusion 120 kcal.

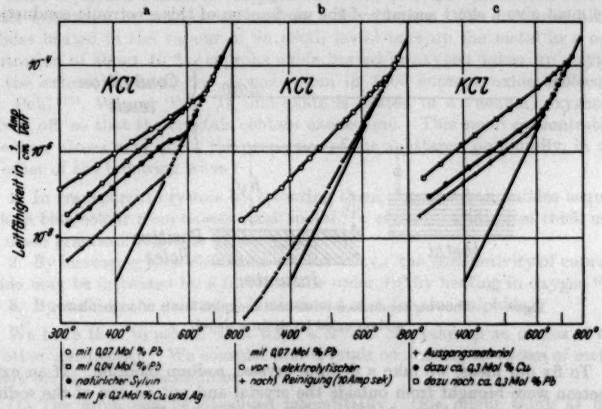


Figure 4. Ionic conductivity of potassium chloride.
The curves are for specimens with different contents of impurity.

(cf. Barrer [13]). Ionic crystals also show a structure-sensitive ionic conductivity. Figure 4 shows some results for potassium chloride (from Lehfeldt [14]); the results may be represented by a formula of the type

$$\sigma = \sigma_0 e^{-Q_0/kT} + \sigma_1 e^{-Q_1/kT},$$

where σ_0, Q_0 are structure-sensitive and $\sigma_0 \ll \sigma_1$, $Q_0 < Q_1$. The structure-sensitive term may be due to diffusion along the boundaries of the crystallites of the mosaic structure; it may also be due to certain kinds of impurities in solid solution. Koch and Wagner[15] have measured the conductivities of crystals of AgCl and AgBr containing small quantities of $CdCl_2$ and $CdBr_2$ in solid

solution. In these mixed crystals, each Cd^{++} ion replaces *two* Ag^{+} ions in the lattice. Thus in the crystal at low temperatures, there will be a number of vacant lattice points equal to the number of cadmium ions. These will make a contribution to the conductivity of the form

$$\sigma = Ae^{-U/kT},$$

where U is the activation energy for the motion of a hole, and A depends on the concentration of cadmium. Koch and Wagner suggest that the structure-sensitive term is often of this type.

§3. MOTION OF ELECTRONS IN POLAR CRYSTALS

At sufficiently low temperatures, and in the absence of illumination, all polar crystals are insulators; under illumination with suitable wave-lengths, or at sufficiently high temperatures, many such substances show a small *electronic* conductivity. In order to explain reactions in these substances, we must give a short account of the mechanism of this electronic conduction.

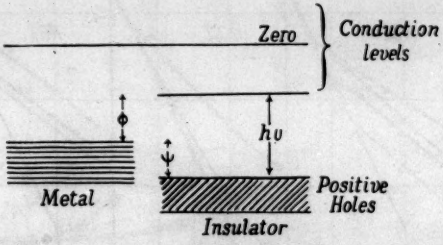


Figure 5. Showing conduction levels and occupied levels of an insulator, and occupied levels for a metal.

To fix our ideas, we take a definite crystal, sodium chloride. If an extra electron were brought from outside the crystal and put on one of the sodium ions, it might be said to form a sodium "atom". However, according to deductions from quantum mechanics, this electron would not stay on any particular ion, but would move easily from one ion to the next, no activation energy being necessary for the process. The possible energies for such an electron would have a lowest value (denoted by χ) for the electron at rest, and any higher value, depending on the kinetic energy of motion of the electron through the crystal. It is customary to call this band of energy levels the "conduction band". Electrons may be raised into the conduction band either by the absorption of radiation or by thermal activation.

The presence of electrons in the conduction band is not the only possible cause of electronic conductivity. If an electron is removed from one of the chlorine ions in sodium chloride, a chlorine atom is formed. According again

to quantum mechanics, an adjacent chloride ion can hand over an electron to the atom ; the process can be repeated, and so the place where the electron is missing can move through the crystal. A transport of charge can thus take place.' A place where an electron is missing is called in the literature a " positive hole ". Positive holes behave in many ways like positive carriers of electricity. Their energy will lie in a band of possible energy levels, as shown in figure 5.

The most direct way of determining whether the current in a semi-conductor or in a photo-conducting material is carried by electrons or by positive holes is by a measurement of the Hall constant. In this way, for instance, Engelhardt [16] has shown that the current in cuprous oxide is carried by positive holes.

§ 4. IONIC CRYSTALS OF NON-STOICHIOMETRIC COMPOSITION

Ionic crystals if heated in the presence of one or other constituent take up small quantities of that constituent in solid solution. For instance, alkali halides heated in the vapour of an alkali metal take up the metal in a concentration of about 10^{-4} ; cuprous oxide heated in oxygen takes up oxygen to the extent of about one oxygen atom in 1000 cuprous oxide molecules (cf. Pohl [17], Wagner [18]). If zinc oxide is heated in a vacuum, oxygen is driven off, so that the crystals contain excess zinc. This small concentration of excess atoms may affect the properties of the substance profoundly, in one or other of the following ways :

1. In transparent crystals by colouring them ; most alkali halides acquire a deep blue colour from excess alkali metal ; a crystal a millimetre thick may be made practically opaque in this way.

2. By increasing their electronic conductivity ; the conductivity of cuprous oxide may be increased by a factor of the order 10^7 by heating in oxygen [18].

3. By making the substance luminescent, e.g. for zinc sulphide.

We have then to ask in what ways a crystal can take up an excess of one or other constituent. We consider first crystals containing an excess of metal. There are then two possibilities :

(1) The atom can be dissolved interstitially. One might think at first sight that a zinc atom would be too large a particle to fit into the interstitial positions in the zinc oxide lattice ; but as a matter of fact it does not have to do so ; only the zinc *ion* has to fit in. The electron is really in the conduction band of the crystal, but is held in position by the field round the zinc ion. Its energy level will be a little below the surface of the conduction band. Its wave function (orbit) may extend over a bigger radius than that of the free atom. It should be emphasized that its absorption spectrum will bear no resemblance whatever to that of the free atom. Whereas in the free atom the electron is attracted to the ion by the force e^2/r^2, in the crystal the force will be reduced to $e^2/\kappa r^2$; κ may be quite large, of the order 10.

(2) There may be a number of lattice points from which the negative ion is missing, and at which an electron is located instead.* Such a crystal will clearly contain excess metal.

Either mechanism provides the crystal with loosely bound electrons, which can absorb light quanta of lower energy than the electrons of the ions in the perfect crystal (thus colouring otherwise transparent crystals), and which can in certain cases be released by thermal agitation, thus giving the crystal an electrical conductivity which increases with the temperature.

An electro-negative constituent, e.g. oxygen in an oxide, can be absorbed by the lattice in the following way :

(3) There will be a number of lattice points from which a positive ion is missing, and one of the ions adjacent to this vacant lattice point will have an electron missing. In other words, we have positive holes trapped in the neighbourhood of vacant lattice points.

We shall now review a few of the properties of particular crystals of non-stoichiometric composition.

(a) *Alkali halides with excess metal.* As already mentioned, alkali halide crystals heated in alkali vapour acquire a deep colour, a characteristic bell-shaped absorption band appearing in the visible. The properties of such

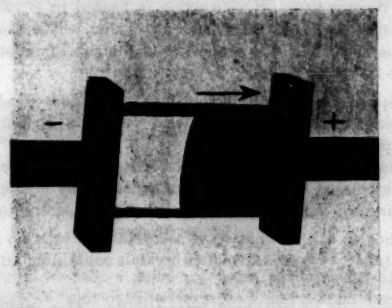


Figure 6. Migration of F-centres in a field (from Pohl [17]).

crystals have been investigated in detail by R. W. Pohl and his school,† who call the absorbed alkali atoms *Farbzentren* (F-centres or colour centres). For reasons which cannot be discussed here, it is practically certain that the excess metal is absorbed by process (2) above : an F-centre is thus an electron trapped at a lattice point from which the negative ion is missing. The identical absorption band, but with a much weaker intensity, can be obtained by irradiating

* The electron may be thought of as located on one of the adjacent positive ions ; it will continually jump from one to another.

† For a summary, cf. Pohl [17].

the crystal with x rays ; the explanation is that electrons ejected from x-ray levels are trapped at the vacant lattice points of which, as we have seen, there will be a certain number frozen-in at room temperature.

Some experiments carried out in Pohl's laboratory give important information about the mechanism of the diffusion of colour centres. The experiment is illustrated in figure 6 ; a crystal containing F-centres is mounted between two electrodes ; at temperatures above about 200° C. the colour centres can be observed to migrate towards the anode in a swarm. The mobility v may thus be determined. It is not immediately obvious why the F-centres move at all in a field, since an F-centre consists of an electron replacing a negative ion in the lattice, and the electron can only *change places* with negative ions. The explanation is that the F-centres at these temperatures have partially dissociated ; every now and then an electron escapes from the neighbourhood of a vacant lattice point into the conduction band of the crystal, and it will then drift towards the anode until it finds another vacant lattice point. There are a large number of vacant lattice points present, because the crystal is an ionic conductor.

If no electric field is present, a coloured strip of the crystal will broaden, diffusion taking place. Pohl has found that the diffusion coefficient D and the mobility v of the F-centres are connected by the relation (cf. equation (4))

$$D/v = kT/e. \qquad \ldots.(5)$$

This shows that F-centres diffuse only after dissociation ; the electron and the vacant lattice point move through the crystal separately.

(b) *Zinc oxide and sulphide.* Zinc oxide is a semi-conductor whose conductivity increases on heating in a vacuum, when oxygen is driven off. It is believed that the excess zinc is held in interstitial positions. These zinc atoms easily lose an electron when the temperature is raised ; the sign of the Hall coefficient shows that the current is carried by electrons, not by positive holes[19].

In zinc sulphide phosphors, it has been suggested by Seitz [20] that the luminescent centres are interstitial metal atoms, either of zinc or of the activating metal.

(c) *Cuprous oxide.* This substance is a semi-conductor of which the conductivity increases on heating in oxygen, so that excess oxygen is taken up. The excess oxygen, which may be present up to concentrations of about one part in 1000, is believed to be included in the crystals by process (3) above. Measurements of the Hall coefficient [16] show that the current is carried by positive holes.

Wagner [18] and his collaborators have measured the concentration of oxygen and the electrical conductivity of cuprous oxide in equilibrium with oxygen at 1000° C. as a function of the pressure p of oxygen. Each oxygen molecule will give rise to two O^{--} ions in the crystal, and hence to four lattice points from which a Cu^{+} ion is missing, and to four " positive holes ". If the temperature is high enough for the " holes " to have left the vacant lattice

points at which they are normally trapped, we expect the oxygen concentration c and hence the conductivity σ to vary as $p^{1/8}$. Wagner in fact found that

$$\sigma \propto p^{1/7},$$
$$c \propto p^{1/5}.$$

The discrepancy between theory and experiment is presumably due to incomplete dissociation.

§ 5. FORMATION OF A NEW PHASE IN SOLID REACTIONS. OXIDATION; TARNISHING REACTIONS

We are now in a position to turn to the proper subject of this article, namely, solid reactions. We shall begin by considering the formation of an oxide or similar layer on a metal surface, for instance, the layer of cuprous oxide formed on copper when exposed to air, or the layer of silver sulphide formed on silver in contact with liquid sulphur. Such reactions have been called by Wagner *Anlaufvorgänge*, or " tarnishing reactions "

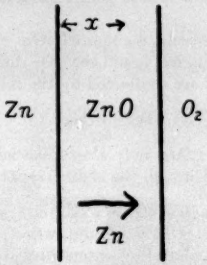


Figure 7. Illustrating a tarnishing reaction. The arrow shows the direction of motion of Zn atoms.

A reaction of this type is shown schematically in figure 7. It is characteristic of such reactions that the two reacting substances, in this case zinc and oxygen, are separated from each other by the product, in this case zinc oxide. Thus either the metal must diffuse through the oxide layer and react with the oxygen at the surface between it and the oxide, or the oxygen must diffuse through the oxide and react with the metal at the interface between it and the oxide. In the majority of cases it is the metal atom which is mobile.

Let us consider in greater detail the case of the oxidation of zinc. Zinc oxide does not absorb excess oxygen in appreciable amounts, but, as we have already seen, it can take up excess zinc. Thus in the oxide layer adjacent to the zinc there will be an equilibrium set up ; the oxide will contain a certain number n_0 of excess zinc atoms per unit volume characteristic of zinc oxide in contact with zinc metal at the temperature considered. At the surface in contact with oxygen, on the other hand, excess zinc atoms may be assumed to react with oxygen as soon as they arrive ; the concentration there will therefore

be zero. There will thus exist in the oxide layer a concentration gradient. If $n(x)$ is the number of excess zinc atoms per unit volume at a distance x from the metal surface, the number N of atoms crossing unit area per unit time is given by

$$N = -D\, dn/dx,$$

where D is the diffusion coefficient. On integrating across an oxide layer of thickness x_0 we obtain

$$Nx_0 = Dn_0.$$

If Ω is the volume of the oxide per metal ion, ΩN is equal to dx_0/dt, the rate of growth of the layer. We thus have

$$\frac{dx_0}{dt} = \frac{Dn_0}{\Omega}\frac{1}{x_0},$$

and thus for the thickness x_0 at a time t after the process starts

$$\tfrac{1}{2}x_0^2 = Dn_0 t/\Omega. \qquad \ldots.(6)$$

Parabolic rates of growth of this type, in which the thickness varies as $t^{1/2}$, are very frequently observed in the formation of oxide films on metal, both in

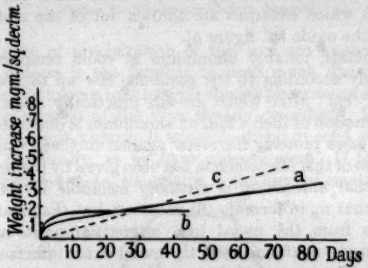


Figure 8. (*a*) Growth of an oxide film on copper. (*b*) Growth of an oxide film on aluminium. (*c*) Square of curve (*a*), showing parabolic law (experimental data from Carpenter and Robertson, *Metals*, **1**, 530, 540).

the case of films formed at room temperature which take years to grow, and those formed at high temperatures which take a few minutes. An example is shown in figure 8 (*a*) and (*c*).

It will be noticed that, for the case of zinc oxide discussed above, the reaction rate is independent of the pressure of oxygen, provided only that the oxygen pressure is sufficient to react with zinc atoms as soon as they arrive at the surface. Wagner [21] in fact has shown that the rate of reaction is independent of pressure for pressures from 1 to 0·01 atmosphere. A recent paper by Vernon and others [41] has shown, however, that the oxidation of zinc does not obey the parabolic law.

The case of the oxidation of copper is different. Cuprous oxide does not take up excess copper, but, as we have seen, it takes up excess oxygen by

forming vacant copper lattice points. The diffusion process taking place consists in a movement of vacant lattice points from the oxide-oxygen interface inwards (actually therefore the copper atoms move !). In this case the quantity n_0 in (6) will refer to the concentration of vacant lattice points at this interface—which is just twice the concentration of excess oxygen atoms. As we have seen, this is proportional to a power of the oxygen pressure, p. Wagner [21] finds that the rate of reaction is proportional to $p^{1/7}$.

Wagner [22] has made the suggestion that the diffusing atoms dissociate into free electrons in the conduction band and interstitial ions or vacant lattice points. The quantity D in formula (6) will then refer to ions or to electrons, whichever restricts the rate of diffusion. In cuprous oxide, for example, it will refer to the diffusion coefficient of the ions ; the substance when of stoichiometric composition is an insulator, and contains no mobile ions, so each electron will have to wait for its own ion to drift right across the layer. In substances which are good electrolytic conductors, however, D refers to the diffusion coefficient of the electrons. For such substances (e.g. silver in contact with silver halides), the metal ions will pass from the metal through the layer of salt or oxide relatively easily ; what limits the rate of reaction is the speed with which electrons are thrown out of the metal into the conduction band of the oxide (cf. figure 5).

On certain metals, notably aluminium at room temperature, the oxide film grows roughly according to the parabolic law up to a certain thickness of the order 10^{-6} cm., after which growth practically stops, unless the film breaks. The formation of such a film on aluminium is illustrated in figure 8 (*b*). The film in such cases protects the metal against further oxidation.

An explanation of this phenomenon has been given by the present author [23]. It is suggested that aluminium is virtually insoluble in the oxide at room temperature, so that n_0 in formula (6) vanishes, but that aluminium ions can nevertheless pass from the metal into interstitial positions in the oxide ; the work function $(\phi-\psi)$ (figure 5) at the metal-oxide interface is, however, too great to allow of any emission of electrons from the metal into the conduction band of the oxide, at room temperature. For very thin films, however, the electrons can be transferred from the metal to the oxide-oxygen interface where the reaction is taking place, without receiving any thermal energy, by means of the quantum-mechanical tunnel effect. It may be shown that an electron can penetrate a " potential barrier " some 50 A. thick in times of less than a day, but for much thicker layers the times required reach astronomical orders of magnitude.

We have here, then, another reaction whose rate is determined by the rate at which electrons can be transferred from one reacting component to the other, rather than by the rate at which ions move. We shall find that this is true also of many other reactions in solids.

§ 6. REACTIONS INVOLVING PRECIPITATION. REDUCTION

Let us consider two solid elements or compounds which are mutually soluble in each other. The phase-equilibrium diagram may be as in figure 9, there being two one-phase regions separated by a two-phase region A+B. If a

solid solution having the composition and temperature represented by the point P is cooled suddenly to the point Q, so that the solution becomes supersaturated, precipitates having the composition B will form either in or on the surface of the original solid.

The conditions for the formation of a nucleus on which precipitation can begin will be discussed in §8; here we shall discuss only the process of precipitation, once a nucleus has been formed.

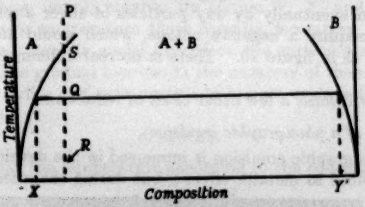


Figure 9. Phase equilibrium diagram.

Very many cases of precipitation of this sort are known among metallic alloys, and the process is too familiar to need much discussion here. It is perhaps less generally known that exactly similar phenomena occur for solutions of metals in ionic crystals of the type already described. If an alkali-halide crystal is heated in alkali vapour, it takes up metal in solid solution, the metal being dispersed atomically, as we have already seen. If the crystal is then cooled *slowly*, the excess metal coagulates in the form of small metallic particles

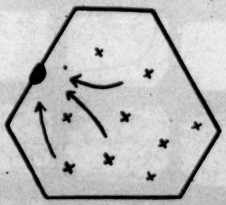


Figure 10. Showing the formation of a silver speck in a halide grain.

in the interior of the crystal; these give quite a different absorption spectrum from the F-centres.

According to a theory of the process given by Gurney and the present author [24], the photo-chemical reduction of silver halides (the print-out process in photographic emulsions) is of a somewhat similar nature. Silver halides are ionic conductors at room temperature, and the conductivity is due to the motion of silver ions in interstitial positions (the Frenkel mechanism, §2). They also show photo-conductivity, the absorption of light raising electrons

to the conduction band, where they are free to move. Before illumination, therefore, a silver-halide crystal contains interstitial ions; after illumination it contains as well a much smaller number of free electrons. The solubility of silver in its chloride or bromide is probably negligibly small, so that we have in the crystal a *supersaturated* solution of silver; this will condense on any silver nucleus present.

In their original paper Gurney and Mott gave a rather more dynamical picture of the phenomenon; they pictured the electrons released by the light as being trapped eventually by any particles of silver already there. These would thereby acquire a negative charge, which would attract the positive interstitial ions, as in figure 10. There is no real difference between the two explanations.

We shall now discuss a few other cases of reduction.

(a) *Development of a photographic emulsion*

When a photographic emulsion is immersed in the developer, the grains of bromide are reduced to metallic silver; but during the intervals of time used

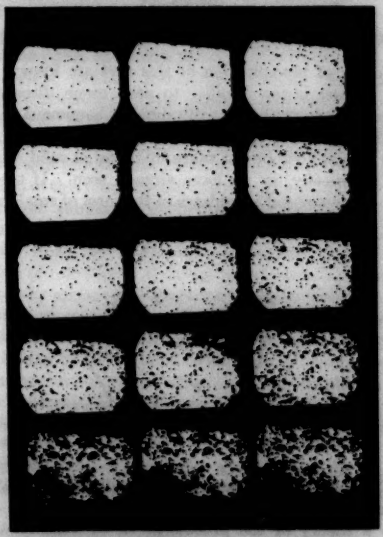


Figure 11. Metallic nuclei formed on barium azide crystals with increasing time of heating.

in practice only those grains are attacked which already contain a speck of metallic silver (the latent image). According to Gurney and Mott the process

takes place as follows: the molecules of the developer hand over electrons to the metallic silver speck, and raise its potential relative to that of the halide grain. The interstitial silver ions which are present in the halide are then attracted to the silver speck; they will move up to it and adhere to it. Of course, as fast as the silver ions in the grain are used up, new ones are formed.*

(b) *Reduction of metallic azides by heat; autocatalytic reactions*

The reduction of metallic azides, such as KN_3, $Na(N_3)_3$ by heat has been investigated in some detail, notably by Garner [25], [26] and his co-workers. If kept at a suitable temperature these crystals decompose slowly, nitrogen being given off and metallic nuclei formed on the surface and in the interior of the crystals. The gradual increase in the numbers of these nuclei is shown in figure 11; the radius r of each nucleus in barium azide is found to increase at a constant rate:

$$dr/dt = A,$$

where A varies with the temperature according to the equation

$$A = A_0 e^{-23,000/RT}$$

The total amount of nitrogen given off is found to increase at first slowly, during the so-called induction period, during which few, if any, metallic nuclei appear, and then much more rapidly; the rapid increase is associated with the formation and growth of nuclei. This is illustrated in figure 12.

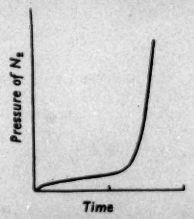


Figure 12. Showing the increase of nitrogen pressure in the decomposition of azides.

The interpretation of the phenomenon given by the present author [27] is as follows: Initially, nitrogen is driven off from the surface, in just the same way that oxygen is driven, for instance, out of zinc oxide when it is heat-treated to increase its conductivity. The metal atoms left behind diffuse into the interior of the crystal, as interstitial atoms of the type already described. When the concentration of interstitial atoms reaches a certain value, metallic nuclei will be formed. The conditions under which these nuclei can be formed will be discussed in the next section.

* We must add that this picture of development is by no means universally accepted; in any case it applies only to the so-called chemical development, not to physical development, where the solution contains silver ions which are deposited on the latent image.

Once a nucleus is formed, however, it grows much more rapidly than can be accounted for by this mechanism—giving rise to the accelerated emission of nitrogen already mentioned. This autocatalytic growth may be accounted for as follows: We may assume that at these temperatures the crystals are good ionic conductors; and thus that a considerable number of metal ions will have left their normal positions and will be in interstitial positions. These ions can move through the crystal, as in figure 10, and adhere to the metal, if the compensating electrons can be provided from somewhere. These electrons can only be supplied from the azide ions of the crystal which are in contact with the metal; in other words, from the electrons in the full band of energy levels of the azide crystal. As figure 5 shows, the energy ψ required to eject an electron from this full band into the metal is less than the quantum energy $h\nu$ required to produce an electron (and thus photo-chemical reduction) in the body of the azide crystal. The energy ψ is to be identified with the observed energy of activation of 23 kcal ($\sim$1 ev.). Autocatalytic reduction of this type will only occur for substances where this energy ψ is abnormally small, so that thermal agitation is sufficient to produce the required electrons.

The "positive hole" formed when an electron is moved from the full band to the metal, will diffuse to the surface. An azide radicle on the surface will presumably break up, giving nitrogen.

§7. INFLUENCE OF STRAINS ON PRECIPITATION

We have now considered three processes in which precipitation of metal occurs: photolytic reduction of silver halides, development of silver halide emulsions and decomposition of metallic azides. In all these processes, if our explanations are correct, the halogen or nitrogen escapes from the surface

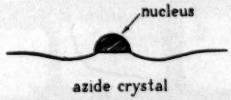


Figure 13. Type of metallic nucleus to be expected in decomposition of azides.

of the crystal, and *not* from the metal-compound interface. Interstitial ions, which may have come from quite distant parts of the crystal, flow through the crystal and join the metal at the interface; but nothing is removed from the interface to make room for the growing nucleus. We expect then that a precipitate can only form at an external surface, or at the surface of an internal crack, and that it will be pushed outwards from the surface, and appear as in figure 13; within the solid a crystal could not grow because there would be no room for it.

Wischin [28] has found microscopic evidence that the metallic nuclei found on barium azide protrude from the crystal surface. Again, photographs due to Rabinowitsch [28] of the development of silver halide grains show a similar phenomenon.

We return now to the consideration of metals. Most alloy systems consist of substitutional solid solutions, rather than interstitial solutions. In such systems precipitates can certainly form within single crystals—not only at intercrystalline boundaries. If now in metallic alloys diffusion consists of a

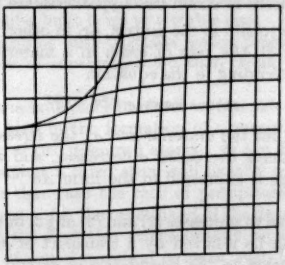


Figure 14. Strain set up in an isotropic solid by a spherical occlusion.

simple interchange of atoms—the first process considered in §2—then the number of atoms moving into the hole occupied by the precipitate is equal to the number moving out. Thus if the volume per atom of the precipitated phase is not exactly equal to that of the alloy from which the precipitate is formed (the matrix), a state of strain will be set up in the neighbourhood of the precipitate. Suppose, for example, that the volume per atom of the precipi-

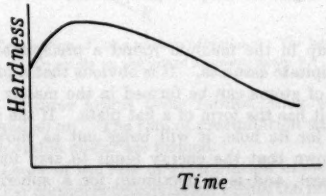


Figure 15. Typical hardness-time curve.

tate is the larger of the two ; then the precipitate will be in a hole where there is not room for it ; it will therefore be under compression. The type of strain set up by a spherical occlusion is shown in figure 14.

These strains are of great importance theoretically, because to them is supposed to be due the hardness acquired by certain alloys on ageing (age hardening). The yield point of such an alloy, according to an extension of

G. I. Taylor's [29] theory of the yield point, due to the present author [30], is of the order of magnitude of the mean internal stress.

After prolonged ageing, alloys in which precipitation is taking place soften again; a typical hardness-time curve is shown in figure 15. This softening must be due to a release of strain round the precipitates. As regards the mechanism of this release of strain two hypotheses are possible:

(1) It may be due to slip in the matrix, or, in other words, to creep. Now according to Orowan [31] the rate of creep in a material varies with stress σ and temperature T according to the equation

$$\text{rate} = \text{const}.\, e^{-\alpha(\sigma_m - \sigma)^2/kT},$$

where α is a constant for the material and σ_m the stress required to cause slip at $T=0$. As the hardness decreases, σ decreases, and we should expect that, if the rate of softening is expressed in the form $Ae^{-Q/RT}$, Q will increase with time.

(2) If diffusion is due to processes (2) and (3) of § 2, but not if it is due to process (1), the strain can be relieved by a transport of matter from the neighbourhood of the occlusion to the boundaries of crystal grains. If this is the correct mechanism, Q should be independent of time.

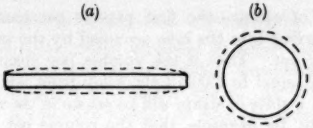


Figure 16. Shapes of precipitates in a metal.

The strain set up in the medium round a precipitate will influence the shape that the precipitate assumes. It is obvious that a precipitate consisting of a given number of atoms can be formed in the matrix with the minimum of strain energy, if it has the form of a flat plate. If the material of the precipitate is too big for its hole, it will bulge out as shown in figure 16 (*a*). Nabarro [32] has shown that the energy tends to zero for vanishingly small thickness of the sheet, and is a maximum for a spherical nucleus (figure 16 (*b*)).

The shape that a precipitate assumes in practice will depend on the degree of supersaturation. If this is high, so that deposition occurs much more rapidly than re-evaporation, we should expect a roughly spherical form. If the degree of supersaturation is low, so that atoms evaporate almost as quickly as they are deposited, we should expect the precipitates to take up the more or less plate-like shape for which the energy is lower. In this way may perhaps be explained the well-known *Widmannstätten* patterns often obtained in alloy systems when one phase crystallizes from another (figure 21).

§ 8. NUCLEAR FORMATION

In water vapour it is well known that condensation does not begin until a certain degree of supersaturation is reached, because the vapour pressure of a very small drop is greater than that of the liquid in bulk, owing to the surface tension. Thus if a small drop were formed by the accidental coming together of a number of molecules, it would usually evaporate before the arrival of further molecules built it up to a stable size. Further, the presence in the atmosphere of ions or dust particles makes possible the formation of drops at a lower degree of supersaturation than would be possible in their absence.

Analogous phenomena take place in reactions in solids in which precipitates are formed. The first step in the formation of a precipitate will in all cases be the coming together of two dissolved atoms. If the degree of supersaturation is high, it may be that a third atom will usually come along before the first two separate. In that case the rate of precipitation will be controlled by the diffusion coefficient. On the other hand, if the degree of supersaturation is low, it may be that the first two atoms usually separate before the third comes; if the degree of supersaturation is lower still, then even in the rare

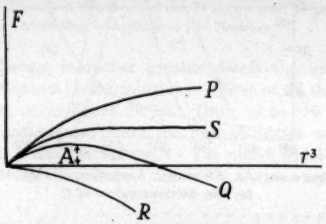


Figure 17. Free energy, F, of a nucleus of radius r. The lettering refers to the points on the phase diagram, figure 9.

event of a nucleus of three atoms being formed, they will separate before the fourth comes; and so on. The rate of precipitation will then be governed by the rate at which stable nuclei are formed.

A quantitative theory of the rate of precipitation in metals has been given by Becker [33]. Becker considers precipitation within the perfect crystal; there is no question in this theory of the presence of favoured spots for nuclear formation. He works out, using a simple model for the interatomic forces, the free-energy change when a spherical nucleus of radius r is formed by precipitation within the crystal. The type of result which he obtains is shown in figure 17, where the curves are for alloys of the same composition.

If the solution is not supersaturated, the free energy will always increase, as shown by the curve marked P, corresponding to the point P on figure 9. S shows the curve for a composition and temperature on the boundary of the two-phase region, and Q for a temperature just below; the free energy first

rises, reaches a maximum, A, and then falls, as the chemical energy gained through the precipitation, varying as r^3, outweighs the surface energy r^2. For a still lower temperature, at which even the smallest nuclei are stable, one will obtain a curve such as R.

In figure 18 we show the value A of the free energy that must be *gained* before a stable nucleus can be formed, plotted against temperature, for the case of Au and Pt, worked out by Becker. According to him, the rate of nuclear formation is proportional to $De^{-A/kT}$, where D is the diffusion coefficient. Since D varies as $D_0e^{-Q/kT}$, where D_0 and Q are almost independent of temperature, we see that the rate of nuclear formation varies as $e^{-(A+Q)/kT}$. This function plotted against temperature should appear as in figure 18 (*b*).

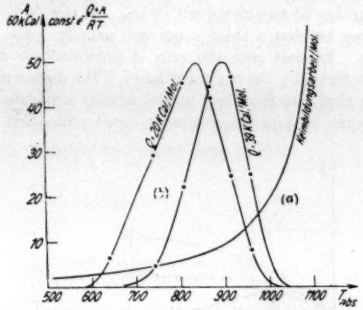


Figure 18. (*a*) Free energy, A, for nuclear formation, and (*b*) rate of nuclear formation, for two different values of Q.

Thus at low temperatures precipitation is slow because diffusion is slow, at high temperatures because of the difficulty of forming a nucleus.

Borelius [34] has treated the same problem from a rather different point of view. Without any specific discussion of nuclear formation, he asks whether a given supersaturated solid solution can separate into two phases of *slightly* different composition without an increase in the free energy. For temperatures and compositions for which this is not the case, precipitation will take a long time to begin.

Some experimental results quoted by Becker in support of his theory are shown in figure 19, where the electrical resistance of gold-platinum alloys is shown, with 30 per cent of platinum. The phase diagram is of the type shown in figure 9. The curve *abcd* represents the resistance for an alloy in equilibrium—i.e. after prolonged annealing. In the range *abc* we have two-phase alloys, from *c* to *d* an alloy of one phase. *a'c* represents the resistance of the quenched alloy—i.e. of the single-phase alloy. The other curves represent the resistance of an alloy which has been quenched to the temperature in question and there annealed for 1, 5, 10, 20 hours. The fall in the resistance

is thus a measure of the precipitation which has taken place ; it will be seen, as demanded by the theory, that there is an optimum temperature for precipitation.

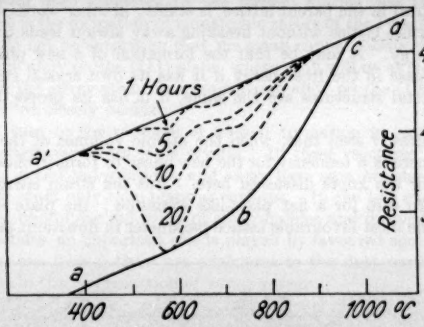


Figure 19. Electrical resistance of Au-Pt alloys with 30 per cent Pt, according to Johansson and Hagsten [35].

We must now discuss in rather greater detail the nature of the nuclei which form in substitutional solid solutions. First of all the nuclei must form *in* the parent lattice, as in figure 20 (*a*). Only when the nucleus assumes a certain size will it break away from the parent lattice and assume its own crystal structure and lattice parameter.* The nuclei which form in the parent lattice are called in the literature "knots". As examples we may quote

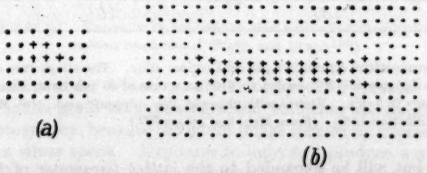


Figure 20. Showing the formation of a "knot" in a crystal. (*a*) shows the initial stages in knot formation; (*b*) a sheet of atoms which have not broken away from the lattice.

Bradley's [36] work on FeNiAl. Although the material in the knots does not form a new phase in the strict sense of the word, there may exist in the alloy, round any one knot before it has grown sufficiently to break away, a thermodynamical equilibrium between the material in the knot and the surrounding solid solution. Since the free energy is always higher for the knot than for the ultimate phase which is going to be precipitated, we may always

* The lattice parameters may be displaced considerably by the strains already discussed.

expect *more* material to remain in solid solution at the knot stage than when the new phase ultimately forms.

A possibility that we have to consider is that in certain alloys no knot, or nucleus formed in the parent lattice, is stable; in other words, that segregation in the perfect lattice without breaking away always leads to an increase in the free energy. It may be that the formation of a new phase will only lead to a decrease in the free energy if it has its own crystal structure; or, if the two crystal structures are the same, if it has its proper lattice parameter.

We have already seen that, when the atomic volumes of the two phases are unequal, there is a tendency for the new phase to form in sheets. This is equally true for the knots discussed here. But the strain energy of a knot will not tend to zero for a flat plate-like structure; the plate can contract or expand to the most favourable lattice parameter in directions perpendicular

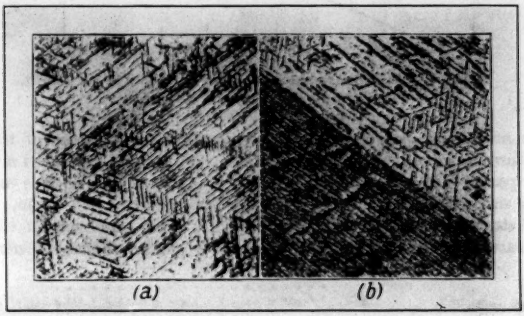


Figure 21. *Widmannstätten* figures in 4 per cent copper alloy. The alloys were annealed at 805°, cooled in furnace to 720°, cooled to 610° at a rate of 5° per hour, then cooled to room temperature in 5 min. Etched with chromic plus sulphuric acid. (*a*) Magnification 300. (*b*) Magnification 1000. (From Mehl and others [38].)

to its plane, but will be expanded to the lattice parameter of the matrix in other directions. It will be under shear, but not under compression (figure 20 (*b*)).

Perhaps the results of Preston [37] on aluminium-copper alloys are to be explained along these lines; he finds that in single crystals aged at room temperature, copper atoms segregate out in sheets along (100) planes, the sheets being not more than one or two atoms thick. That something more than macroscopic strains must be considered here, however, is shown by the fact that segregation occurs on definite crystal planes.

Mehl [38] has obtained *Widmannstätten* formations in copper-silver alloys by slow cooling (cf. figure 21). He finds that islands of copper in silver crystals

are formed parallel to the (111) planes, islands of silver in copper parallel to the (100) planes. These islands are of course large crystals, visible under the microscope, which have long since broken away from the original lattice. It has however been suggested by the present author that the *initial* stage in the process is the formation of monatomic sheets of the type found in Cu–Al. Then it would not be surprising that the small copper atoms should segregate in a silver-rich alloy on the (111) plane, where they will be as closely packed as possible, and the silver atoms in a copper-rich alloy on the (100) plane, where they are not too closely packed.

We now turn to the problem of nuclear formation in reactions in which metal is precipitated in an ionic crystal. In these reactions we are usually dealing with interstitial solutions, and as we have seen, we should expect that precipitation could only take place at surfaces. In the two examples that we shall discuss, photolytic reduction of silver halides and thermal decomposition of metallic azides, an important rôle is played by favoured spots on the surface where nuclei can form ; these are analogous to the dust particles or gaseous ions which help the condensation of water vapour.

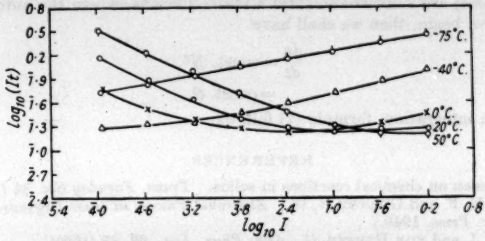


Figure 22. Exposure It required to give a specified density for various intensities I (Webb and Evans [40]).

The conditions under which a nucleus or speck of metallic silver can be formed in a halide grain of a photographic emulsion are important for the science of photography, because the latent image formed by exposure is believed to consist of a silver speck. Exposure to light will produce a number of free electrons in the conduction band of the grain ; but if these are not used up quickly in forming a stable metallic nucleus, they may recombine with the halogen atoms from which they came, and be wasted.

As the work of Sheppard [39] has shown, the sensitivity of a photographic emulsion depends very largely on the presence on the surface of the grains of minute specks of silver sulphide. Gurney and Mott [24] suggested that the function of these " sensitivity specks " is to trap electrons, so that they will stay in one place long enough for silver ions to move up to them.

Just as in alloys, a small precipitate may be unstable if the temperature is not too low. This is shown by the falling off in the sensitivity of most

photographic emulsions for exposures in which the intensity of light is very weak (cf., for example, Webb and Evans [40]). An exposure to light of low intensity will produce a low concentration of free electrons in the grain. Thus if one electron is trapped on a sensitivity speck, and has attracted one ion, it may escape before the second or third electrons arrive to build up a stable nucleus. As we should expect, the effect becomes less marked as the temperature is lowered, and at −75° C. does not appear for the longest exposures used.

We turn now to the formation of metallic nuclei on crystals of metallic azides when they are kept at an elevated temperature. Wischin [26] has shown that on crystals of barium azide at 100° C. the number n of nuclei per unit area of surface shown in figure 11 depends on the time according to the law

$$n = At^3. \qquad \ldots.(7)$$

The constant A varies from one crystal to another, and must depend on the number of favourable spots for nuclear formation. The variation with time has been interpreted in the following way [27]:—We have seen that there is a steady emission of nitrogen from the surface, so the concentration N of electrons increases linearly with the time. If then *two* electrons (with either one or two ions ?) are sufficient to form a stable nucleus, where the autocatalytic reaction can begin, then we shall have

$$\frac{dn}{dt} = \text{const. } N^2$$
$$= \text{const. } t^2,$$

whence, on integration, formula (7) follows.

REFERENCES

(1) Discussion on chemical reactions in solids. *Trans. Faraday Soc.* **34** (1938).
(2) Mott, N. F. and Gurney, R. W. *Electronic Process in Ionic Crystals.* (Oxford Univ. Press, 1940.)
(3) Groh, J. and von Hevesy, G. *Ann. Phys., Lpz.*, **63**, 85 (1920).
(4) Steigman, J., Shockley, W. and Nix, F. C. *Phys. Rev.* **56**, 13 (1939).
(5) von Hevesy, G. and Seith, W. *Z. Phys.* **56**, 790 (1929); **57**, 869 (1929).
(6) Frenkel, J. *Z. Phys.* **35**, 652 (1926).
(7) Wagner, C. and Schottky, W. *Z. phys. Chem.* B, **11**, 163 (1930).
(8) Lennard-Jones, J. E. and Devonshire, A. F. *Proc. Roy. Soc.* A, **170**, 464 (1939).
(9) Tubandt, C. and Seith, W. *Z. Phys.* **57**, 869 (1929).
(10) Jost, W. *J. Chem. Phys.* **1**, 466 (1933).
(11) Mott, N. F. and Littleton, M. J. *Trans. Faraday Soc.* **34**, 485 (1938).
(12) Rhines and Mehl. *Trans. Amer. Inst. Min. Met. Engrs.* **128**, 185 (1938).
(13) Barrer, R. M. *Proc. Phys. Soc.* **52**, 58 (1940).
(14) Lehfeldt. *Z. Phys.* **85**, 717 (1933).
(15) Koch, E. and Wagner, C. *Z. phys. Chem.* B, **38**, 295 (1937).
(16) Engelhard, E. *Ann. Phys., Lpz.*, **17**, 501 (1933).
(17) Pohl, R. W. *Proc. Phys. Soc.* **49** (extra part), 1 (1937).
(18) Gundermann, J. and Wagner, C. *Z. phys. Chem.* B, **37**, 155 (1937).
(19) Fritsch, O. *Ann. Phys., Lpz.*, **22**, 375 (1935).
(20) Seitz, F. *J. Chem. Phys.* **6**, 454 (1938).
(21) Wagner, C. and Grunewald, K. *Z. phys. Chem.* B, **40**, 455 (1938).
(22) Wagner, C. *Z. phys. Chem.* B, **21**, 25 (1933).

(23) MOTT, N. F. *Trans. Faraday Soc.* **35**, 1175 (1939).
(24) GURNEY, R. W. and MOTT, N. F. *Proc. Roy. Soc.* A, **169**, 151 (1938)
(25) GARNER, W. E. and MAGGS. *Proc. Roy. Soc.* A, **172**, 299 (1939).
(26) WISCHIN, A. *Proc. Roy. Soc.* A, **172**, 314 (1939).
(27) MOTT, N. F. *Proc. Roy. Soc.* A, **172**, 325 (1939).
(28) RABINOWITSCH, A. *Trans. Faraday Soc.* **34**, 920 (1938).
(29) TAYLOR, G. I. *Proc. Roy. Soc.* A, **145**, 362 (1934).
(30) MOTT, N. F. and NABARRO, F. *Proc. Phys. Soc.* **52**, 86 (1940).
(31) OROWAN, E. *Z. Phys.*, **39**, 605 (1934).
(32) NABARRO, F. *Proc. Phys. Soc.* **52**, 90 (1940).
(33) BECKER, R. *Ann. Phys., Lpz.*, **32**, 128 (1938).
(34) BORELIUS, G. *Ann. Phys., Lpz.*, **28**, 507 (1936).
(35) JOHANSSON, C. K. and HAGSTEN, G. *Ann. Phys., Lpz.*, **28**, 520 (1937).
(36) BRADLEY, A. J. *Proc. Phys. Soc.* **52**, 80 (1940).
(37) PRESTON, G. D. *Proc. Roy. Soc.* A, **167**, 526 (1938).
(38) BARRETT, C. S., KAISER, H. F. and MEHL, R. F. *Trans. Amer. Inst. Min. Met. Engrs.* no. 595 (1933).
(39) SHEPPARD, S. E., TRIVELLI, A. P. H. and LOVELAND, R. P. *J. Franklin Inst.* **200**, 51 (1925).
(40) WEBB, J. H. and EVANS, C. H. *J. Opt. Soc. Amer.* **26**, 367 (1936).
(41) VERNON, W. H. T., AKEROYD, E. I. and STROUD, E. G. *J. Inst. Met.* **6** (1939).

THE THEORY OF MOLECULAR STRUCTURE

By W. G. PENNEY, D.Sc.
Imperial College, London

THIS is the second article on the theory of molecular structure to be given in these *Reports*. The first appeared in vol. II. Several interesting discoveries have been made since then, many of them as the result of lengthy numerical calculations. The *Report* is divided into five main sections, and a final small section which mentions papers connected with, but not lying entirely in, the present field of discussion.

§ 1. ENERGY AND RELATED PROPERTIES BY VARIATION METHODS

Excellent progress has been made since the last *Report* in applying variation methods to obtain the energies of different states of simple polyatomic molecules. The advances which have been made have resulted mainly from lengthy numerical integration, often with mechanical aid.

States of H_2. The variation principle, when used to calculate the energy states of molecules, just like corresponding calculations in other branches of mathematical physics, is powerful only when applied to the most stable or lowest eigen-value. This need not happen if the system has sufficient symmetry. For example, the fundamental vibrations of a string are either symmetrical or anti-symmetrical about the centre. The eigen-functions fall into two similar classes, and for this reason the variation principle could be applied immediately to estimate the gravest fundamental of either type.

The molecule H_2 has great symmetry. Even when rotation and vibration of the molecule are neglected, three symbols are needed to specify the symmetry elements of the Schrödinger function of any energy level. These refer respectively to the resultant electronic spin (singlet or triplet); the even or odd property of the Schrödinger function on reflexion of the electronic co-ordinates in the central plane (g or u); and another symbol Σ, Π, . . ., denoting whether the component of electronic orbital angular momentum about the axis of the molecule, in units $h/2\pi$, is 0, 1, . . .

James, Coolidge and Present [1] have made an important contribution to the theory of molecular structure by calculating potential energy curves and Schrödinger functions of states of H_2. They consider $(1s\sigma)^2\ {}^1\Sigma_g$, $(1s\sigma,\ 2p\sigma)\ {}^1\Sigma_u$, $(1s\sigma,\ 2p\sigma)\ {}^3\Sigma_u$ and $(1s\sigma,\ 2s\sigma)\ {}^3\Sigma_g$, each of which is the most stable of its respective type The first is the ground state; the second is the B state; the third is the Heitler-London repulsive state; and the fourth is the lowest stable excited state. The continuous emission spectrum arises from transitions from the fourth to the third.

The starting point of the theory is the separable property of the Schrödinger equation for H_2^+ in spheroidal coordinates. (Details will be found in the book of Pauling and Wilson [2].) If the electronic repulsion term in the equation for H_2 is neglected, a separable solution is also obtained in these coordinates.

Denote the nuclei by a and b, the electrons by 1 and 2, distances involving one or both electrons by r with appropriate suffixes (e.g. r_{1a}, r_{12}), and the internuclear distance by R, all distances being expressed in atomic units (i.e. in terms of a_H, the first Bohr radius of H).

Introduce spheroidal coordinates

$$\lambda_1=(r_{1a}+r_{1b})/R, \qquad \mu_1=(r_{1a}-r_{1b})/R, \qquad \phi_1,$$
$$\lambda_2=(r_{2a}+r_{2b})/R, \qquad \mu_2=(r_{2a}-r_{2b})/R, \qquad \phi_2,$$

where ϕ_1 and ϕ_2 are the two azimuths.

The states which are to be considered are Σ states, and for these the only combination of ϕ_1 and ϕ_2 which can appear in the Schrödinger function is $(\phi_1-\phi_2)$. Now r_{12} and $1/r_{12}$ may be expressed in terms of λ_1, μ_1, λ_2, μ_2, $\phi_1-\phi_2$. Instead of obtaining the Schrödinger function in terms of λ_1, μ_1, λ_2, μ_2, $\phi_1-\phi_2$ let us try to obtain it in terms of λ_1, μ_1, λ_2, μ_2, ρ, where

$$\rho=2r_{12}/R.$$

As an approximation, assume

$$\psi=\Sigma C_{mnjkp}[e^{-\delta_1\lambda_1-\delta_2\lambda_2}\lambda_1^m\lambda_2^n\mu_1^j\mu_2^k\pm e^{-\delta_2\lambda_1-\delta_1\lambda_2}\lambda_1^n\lambda_2^m\mu_1^k\mu_2^j]\rho^p,$$

where the $-$ sign is to be taken for triplet states and the $+$ sign for singlet states. For u states $(j+k)$ must be taken odd, and for g states even. Otherwise, the summation is over zero or positive integral values of m, n, j, k and p.

The variation function which is to be minimised by choosing the C's is

$$H=\int\psi\,\mathscr{H}\psi\,d\tau/\int\psi^2\,d\tau,$$

where $\mathscr{H}$ is the Hamiltonian operator. The minimum value of H is an upper limit to the energy of the molecule, at the particular value taken for R. The approximation to the potential energy curve is thus a plot of H against R; from the curvature at the minimum the force constant and vibration frequency may be found.

When sufficient terms are taken in the expansion for ψ, the best value of the energy varies little with the choice of δ_1 and δ_2, provided that they lie in an appropriate range. Calculations were made with $\delta_1=1$, $\delta_2=0{\cdot}50$, and $\delta_1=\delta_2=0{\cdot}75$. These choices represent two attempts at making immediate allowance for the repulsion between the electrons. The first is equivalent to saying that one electron moves in the full nuclear field, and is bound very closely; the second electron is further out, and the field in which it moves is that of the nuclei, shielded by the inner electron. The shielding is taken to be complete, i.e. one-half of an electron at each nucleus. The first choice for the δ's is thus very suitable for excited states of H_2, where the two electrons are not equivalent. The second choice is equivalent to assuming that each electron half shields the other, and that the effective nuclear charge acting on both electrons is 0·75.

The maximum number of terms in the expansion for ψ was 13. Most of the labour is concerned with computing the integrals occurring in H. The problem of minimizing the energy, once the integrals are known, is comparatively short, because the convergence is good.

The reader is referred to the original papers for the mathematical details of the evaluation of the energy integrals, and for a discussion of the probable accuracy of the energy values and the Schrödinger functions.

Systems HeH^{++} and HeH^{+}. Coulson and Duncanson [3] have compared the results of some half-dozen approximation methods when applied to HeH^{++} and HeH^{+}. The first system is not stable; its potential energy curve can be calculated as accurately as desired by a rapid approximation method devised by Hassé and Baber [4]. The second system is stable. The dissociation energy is calculated to lie with certainty between 0·6 and 2·1 ev., the internuclear distance is 0·76 A., and the fundamental frequency is 3380 cm^{-1} Beach [5], by a less accurate treatment, had previously obtained 0·76 A., and 2800 cm^{-1}

Systems H_3 and H_3^{+}. The H_3 complex dissociates exothermically (7 kcal./mol.) into H_2+H. The complex is of considerable theoretical interest because it is an intermediate product in the ortho-para hydrogen conversion. The H_3^{+} ion has been found in the discharge tube, although nothing is known of its spectrum.

The Schrödinger equation for a system consisting of three protons and one electron is not separable in any coordinate system. Hence the best that can be done to approximate to the Schrödinger functions of H_3, H_3^{+} and H_3^{++} is to use linear combinations of atomic orbitals, usually hydrogenic 1*s* functions, with arbitrary screening constants. Complications immediately arise because of the non-orthogonality of the atomic orbitals.

Coulson [6] has applied the method of molecular orbitals to show that the most probable configuration of H_3^{+} is that of an equilateral triangle, with side 0·85 A. (This result is not quite correct; see later.) The orbitals, apart from normalizing factors which are approximately $1/\sqrt{3}$, are

$$\Psi=\psi_a+\omega\psi_b+\omega^2\psi_c,$$

where ω is a cube root of unity and ψ_a is an atomic orbital of hydrogen centred on proton *a*. The most stable, and only bonding, orbital is that with $\omega=1$; this can accommodate two electrons. The complex values of ω give a degenerate pair of anti-bonding orbitals. An electron added to H_3^{+} must pass into one of these anti-bonding orbitals, and the effect is so great that the molecule assumes a linear form.

A great improvement on Coulson's work has been described in a series of papers by Hirschfelder, Eyring, Rosen, Weygandt, Diamond and Stevenson [7]. Beginning with the approximation of hydrogenic 1*s* atomic orbitals, they have made a complete calculation for the energy levels of the H_3 system in the neighbourhood of its most stable configuration, a linear form, and of H_3^{+} in the form of an isosceles triangle, arising from atomic 1*s* orbitals of H.

Consider the Schrödinger function for a linear form of H_3. Denote the protons from left to right by *a*, *b*, *c*, and the three electrons by 1, 2, 3, each

with its two possibilities for spin, α and β. The lowest state of the system is a doublet. We need consider only one component of the doublet, say that one with spin component $M_s=\frac{1}{2}$, i.e. an α-state. If polar states are neglected, there are two states with $S=M_s=\frac{1}{2}$. Their Schrödinger functions may be constructed as combinations of ψ_1 and ψ_3, where ψ_1 and ψ_2 are functions representing valence bonds between the electrons on a and b and on b and c respectively.

$$\psi_1=\begin{pmatrix}a\,b\,c\\ \alpha\,\beta\,\alpha\end{pmatrix}-\begin{pmatrix}a\,b\,c\\ \beta\,\alpha\,\alpha\end{pmatrix}, \qquad \psi_2=\begin{pmatrix}a\,b\,c\\ \alpha\,\alpha\,\beta\end{pmatrix}-\begin{pmatrix}a\,b\,c\\ \alpha\,\beta\,\alpha\end{pmatrix},$$

where each of the symbols ( ) represents a Slater determinant. For example

$$\begin{pmatrix}a\,b\,c\\ \alpha\,\beta\,\alpha\end{pmatrix}=\begin{vmatrix}\psi_a(1\alpha) & \psi_b(1\beta) & \psi_c(1\alpha)\\ \psi_a(2\alpha) & \psi_b(2\beta) & \psi_c(2\alpha)\\ \psi_a(3\alpha) & \psi_b(3\beta) & \psi_c(3\alpha)\end{vmatrix}, \quad \psi_a(1\alpha)=Ne^{-pr}\delta(m_{s_1},\tfrac{1}{2}),$$

where N is a numerical factor, p is a parameter to allow for shielding, and δ is a spin delta-function.

There are six polar states of H_3, namely the six permutations of the arrangement two electrons on a, one on b, none on c. These are all necessarily doublets, because if two electrons occupy the same atomic orbital their spins are necessarily coupled to zero.

States with three electrons on one proton are impossible because they are contrary to the Pauli principle.

Define the six polar states by functions $\psi_3, \ldots, \psi_8$, where, for example,

$$\psi_3=\begin{pmatrix}b\,b\,c\\ \alpha\,\beta\,\alpha\end{pmatrix}.$$

The Schrödinger function of the most stable state of H_3 is a linear combination of $\psi_1, \ldots, \psi_8$, chosen to make the energy least, say,

$$\Psi=k_1\psi_1+k_2\psi_2+\ldots+k_8\psi_8.$$

The problem of finding the energy W, in the usual way, is reduced to solving the secular determinant, of eight rows and columns,

$$\mathrm{Det}\{H_{ij}-W\triangle_{ij}\}=0,$$

where

$$H_{ij}=\int\psi_i\mathscr{H}\psi_j\,d\tau,$$

$$\triangle_{ij}=\int\psi_i\psi_j\,d\tau,$$

$$\mathscr{H}=-\tfrac{1}{2}(\nabla_1^2+\nabla_2^2+\nabla_3^2)-(1/r_{a_1}+1/r_{a_2}+\ldots+1/r_{c_3})$$
$$+(1/r_{12}+1/r_{23}+1/r_{31}+1/r_{ab}+1/r_{bc}+1/r_{ac}).$$

Each of the matrix elements H_{ij} is a long series of terms, most of which can be evaluated analytically. The troublesome ones are the three centre integrals, an example of which is

$$L(aa,bc)=\frac{1}{\pi^3}\iint\frac{1}{r_{12}}\exp[-2r_{a_1}-r_{b_2}-r_{b_3}]\,d\tau_1\,d\tau_2.$$

Integration over the coordinates of electron 1 is direct. The remaining

triple integral is best expressed in spheroidal coordinates λ, μ, θ. The θ integration is of the form

$$I(\lambda, \mu) = \int_0^{\pi} \exp[-B(1-A\cos\theta)^{\frac{1}{2}}]\, d\theta,$$

where A and B are functions of λ and μ. The integral I was obtained for over 200 values of the parameters λ and μ with the aid of the differential analyser. Tables of numerical values of integrals in terms of which the three centre integrals may be expressed are given by Hirschfelder and Weygandt [7].

The expression obtained for the energy must be minimized by choosing the effective nuclear charge, and the distances r_{ab}, r_{bc}. The results were $r_{ab}=r_{bc}=0{\cdot}96$ A., compared with $R=0{\cdot}76$ A. in H_2 and $W=3{\cdot}34$ ev., compared with the experimental value 4·13 ev. The calculated activation energy of the ortho-para conversion is thus 25 kcal. compared with 7 observed.

Similar calculations show that the energy of formation of H_3^+ from 2H and H^+ is at least 184 kcal. The shape of the most stable form of H_3^+ is that of an isosceles triangle, with the equal sides, of length 1·79 A., including an angle greater than 60° but less than 90°. The exact dimensions of the triangle are difficult to estimate because the full calculations have to be carried through for various shapes and sizes of the triangle, and the equilibrium position then found by interpolation. The equilateral triangle configuration is definitely not the most stable position.

§ 2. ENERGY AND RELATED PROPERTIES BY PERTURBATION METHODS

The previous section has described calculations which furnish the energy of formation of molecules from atoms directly as the difference of the energy of formation of molecules from electrons and nuclei, and the energy of formation of the separate atoms from electrons and nuclei. This method is at present unmanageable for molecules more complicated than those containing two or three protons and two or three electrons. Progress can only be made by perturbation calculations.

The basic approximation is that the molecular forces do not appreciably distort the orbitals of the atoms. Schrödinger functions for the molecule are thus obtained in terms of the atomic orbitals, and the energy is evaluated as the lowest root of a secular determinant.

The hydro-carbons. The hydro-carbon molecules are interesting to study theoretically by approximate methods because these are good enough to give reasons for the varying directional properties of the four valences of a carbon atom, according to whether the bonds are all single, or only two of them single, or only one of them single, or none of them single. The methods which have been developed permit the details of the structure of even comparatively complex molecules to be calculated. An important simplification in the theory is that polar effects may be neglected, because the bonds are almost completely covalent.

The hydro-carbon molecules fall roughly into two classes. The first can be represented by simple bond diagrams, and the second cannot. Examples

of molecules of the first class are methane, ethane, ethylene and acetylene. The experimental data on such molecules show that the energy of formation can be predicted accurately from a set of single-, double- and triple-bond energies, and that the carbon-carbon internuclear distances for the single bonds is always about 1·54 A., for the double bonds is 1·33 A., and for the triple bonds is 1·20 A. Similarly conserved are the single-, double- and triple-bond force constants. Examples of molecules of the second class are benzene, aromatic carbon compounds containing condensed ring systems, and more generally all molecules containing alternations of single and double bonds. For such molecules the energy of formation observed is greater than that predicted from any single-bond diagram, and the internuclear distances vary over the whole range between the single- and triple-bond distances.

The theories which we shall explain have shown that a molecule for which no single-bond structure can be drawn will behave as if it could be represented by a composite pattern of all of the structures. The most stable state will be more stable than any of those represented by bond structures. Since some of the structures will have double bonds where others have single bonds we have immediately some indication why the internuclear distances can have values between those of the single and double bonds. Molecules of the class we are now discussing are called *mesomeric*, and the quantal effect that produces the extra stability of the molecule over that of any of the bond structures is called *resonance*.

Hückel [8] and Pauling and his collaborators [9] were the first to give a theoretical treatment of mesomerism. They were mainly interested in estimating the extra stability caused by resonance, and compared their results extensively with thermochemical data. Severe limitations were introduced by assuming all carbon-carbon distances equal. Lennard-Jones [10], Penney [11] and others have explained how the theory may be used to predict the distances, the resonance energy and the force constants. Probably the best way of illustrating their arguments is to give a simple example with all of the details.

Butadiene. A suitable molecule is butadiene, of which the chemical formula is $H_2C{=}CH{-}CH{=}CH_2$. The butadiene molecule is planar and all the bond angles are about 120°. Whether the *cis* or the *trans* form is the more stable is uncertain; in either case the two halves of the molecule are similar.

Denote the C atoms successively by a, b, c, d, and use the appropriate suffix or suffixes when it is necessary to distinguish any property associated with a particular C atom or pair of atoms.

The theory sets out to establish that the carbon-carbon internuclear distances $x_{ab}{=}x_{cd}$ and x_{bc} are respectively longer and shorter than the normal double and single bonds, and that the molecule has a greater energy of formation than that predicted from the chemical formula.

The atomic orbitals are of two types; one set is even and the other odd for reflexion in the plane of the molecule. Because of this symmetry the secular equation falls into two parts. One part may be interpreted as giving the energy of the bonds in the plane of the molecule, and the other as the energy of the other bonds. Without further investigation we may assume that the bonds

in the plane are normal, and that their energies may be expressed in terms of the energy of normal single bonds, suitably corrected for compression or extension, as the case may be, from a knowledge of force constants.

Write

$$F=V+E,$$

where V is the energy of the bonds in the plane and E is the energy of the other bonds. Then V and E are functions of the internuclear distances. Assuming $x_{ab}=x_{cd}$, and neglecting the C–H bonds, which are normal,

$$V(x_{ab}, x_{cd})=3V_s^0+k_s(x_{ab}-x_s)^2+\tfrac{1}{2}k_s(x_{bc}-x_s)^2,$$

where V_s^0, k_s, x_s are respectively the energy, force constant and length of a single bond.

The conditions of equilibrium are

$$\partial F/\partial x_{ab}=\partial F/\partial x_{bc}=0.$$

The problem has been reduced to finding E. Let ψ_a, ψ_b, ψ_c, ψ_d be the four odd, or π, orbitals of the C atoms. The expression for each ψ is a $\sigma 2p$ orbital of a C atom, with the axis of quantization perpendicular to the plane of the molecule. At this point there are two methods of approximation, the orbital method and the pair method. The former has been developed mainly by Lennard-Jones, Coulson and collaborators [10], and the second mainly by Penney and collaborators [11]. The results obtained by the two methods are, for practical purposes, identical. Since the two theories stress some interactions at the expense of others, in different but complementary ways, confidence may be placed in the results. The two methods, in fact, produce interpolation formulae which involve only well-known physical constants and properties of the single and double bonds. When experimental values for these latter are put in at the limits of the range of the formulae, values calculated inside the range can hardly be wrong by any measurable quantity.

The orbital approximation. The molecular orbitals for the π electrons fall into two further symmetry classes, two that are even with respect to the centre and two that are odd. Expressions for them are

$$\Psi_1=\alpha(\psi_a+\psi_d)+\beta(\psi_b+\psi_c), \qquad \Psi_3=\gamma(\psi_a-\psi_d)+\delta(\psi_b-\psi_c),$$
$$\Psi_2=-\beta(\psi_a+\psi_d)+\alpha(\psi_b+\psi_c), \qquad \Psi_4=-\delta(\psi_a-\psi_d)+\gamma(\psi_b-\psi_c).$$

Eliminating α and β leads to one quadratic equation, and eliminating γ and δ leads to another:

$$\begin{vmatrix} \epsilon_a-W & \beta_{ab} \\ \beta_{ab} & \epsilon_b+\beta_{bc}-W \end{vmatrix}=0, \qquad \begin{vmatrix} \epsilon_a-W & \beta_{ab} \\ \beta_{ab} & \epsilon_b-\beta_{bc}-W \end{vmatrix}=0,$$

where

$$\beta_{ab}=\int\psi_a(1)\psi_b(1)\{V_a(1)+V_b(1)\}dv_1$$

and $V_a(1)$, for example, is the potential energy of electron 1 in the neighbourhood of atom a. The ϵ are Coulomb integrals, i.e. their integrands are of the form

$\psi^2 V$, but their exact expressions are lengthy because they involve the even orbitals of the bonds in the plane of the molecule. There is no need to give the detailed expressions, for reasons explained later.

The dominant terms in the above equations are the β. Solving these equations as if all the ϵ were zero, and then expanding the roots as power series in the ϵ, we obtain as the energy of the four electrons, two in each of the lowest orbitals,

$$E=2(\epsilon_a+\epsilon_b)-2G[1+2(\epsilon_a-\epsilon_b)^2\beta_{ab}{}^2/G^4+\ldots.], \quad G=\sqrt{(\beta_{bc}{}^2+4\beta_{ab}{}^2)}.$$

Little is known about the magnitudes of the ϵ except that they are small compared with the β. Let us therefore replace $(\epsilon_a+\epsilon_b)$ by a mean value 2ϵ, and neglect second and higher powers of $(\epsilon_a-\epsilon_b)$. Then the function E which we are seeking is

$$E=4\epsilon-2\sqrt{(\beta_{bc}{}^2+4\beta_{ab}{}^2)}.$$

An expression for β as a function of internuclear distance can be obtained as follows. The difference of the energy of the double bond in ethylene and the single bond in ethane is due mainly to the bonding influence of the two π electrons in ethylene, each of energy β. Let $V_d(x)$, $V_s(x)$ be respectively the energies of a double and a single bond, both of length x. Then

$$2\beta(x)=V_d(x)-V_s(x).$$

The functions V_d, V_s may be constructed from the known energies of the bonds and their force constants. Thus

$$V_d(x)=V_d{}^0+\tfrac{1}{2}k_d(x-d)^2,$$
$$V_s(x)=V_s{}^0+\tfrac{1}{2}k_s(x-s)^2,$$

where s and d are the lengths of a single and double bond respectively. Morse functions may be used instead of these parabolic approximations if greater accuracy is desired.

The total energy F of the molecule is now known in terms of the distances, and the equilibrium configuration may be calculated. The results are $x_{ab}=1{\cdot}34$ A., compared with the double-bond distance $1{\cdot}33$ A., $x_{bc}=1{\cdot}43$ A., compared with the single-bond distance $1{\cdot}54$ A. The additional stability of the molecule over that calculated from the chemical formula is about 6 kcal./mol.

Lennard-Jones has carried through calculations similar to those above, to estimate the dimensions and the energies of formation of the benzene ring, the simpler polyenes, the naphthalene molecule and the allyl radical. A summary of the results has been given by Lennard-Jones and Coulson [10]. Penney and Kynch [11] have considered the effect of conjugating a double bond with a benzene ring, and the phenyl-methyl radical. In all cases where comparison with experiment is possible the agreement is excellent.

The pair method. It is possible with the pair method to state at the beginning what approximations are to be made, and from then onwards to make no further approximations. Non-orthogonality of atomic orbitals centred on different atoms, ionic terms, and interactions between non-nearest

neighbours are to be neglected ; perfect pairing of the localized bonds in the plane of the molecule is to be assumed. With these approximations all C—H bonds in planar hydro-carbons are the same. The energy of the molecule apart from the C—H bonds can be expressed in terms of two functions $S(x)$ and $J(x)$. These functions can be constructed from known experimental data ; the equilibrium configuration and the force constants can then be found in the usual way by minimizing the energy.

The energy function for the double bond may be written as the sum of two terms, one a numerical multiple of J (the $\pi-\pi$ exchange integral), and the other a long series of Coulomb and exchange terms, not containing J.

$$V_d(x)=S(x)+\tfrac{1}{2}J(x).$$

The factor $\frac{1}{2}$ arises as follows :—the exchange energy of the paired $\pi-\pi$ electrons is J ; however, there is a similar pair of electrons in the localized trigonal bonds, and they have an exchange interaction $-\frac{1}{2}J$.

Another molecule whose physical constants are well known is benzene. This time we have

$$V_b(x)=S(x)-0{\cdot}066J(x).$$

The second term is built up from $-\frac{1}{2}J$, arising from the electrons in the localized bonds, and $0{\cdot}434\,J$, the average energy per bond of the π-electrons.

An expression for $V_d(x)$ has already been given ; more accurately, Morse functions may be used for V_d and V_b :

$$V(x)=V^0[2e^{-a(x-x_0)}-e^{-2a(x-x_0)}],$$

where V^0, $-2V^0a^2$, x_0 are respectively the energy of the bond, the force constant and the equilibrium distance. Substituting the known constants of the double and benzene bonds, expressions for $S(x)$ and $J(x)$ are obtained.

Consider now the energy of the butadiene molecule. To calculate the energy of the π-electrons one must solve the four-electron problem, well known in activation calculations. The equation to be solved is very simple, being only a quadratic. The method by which the equation is constructed is, however, complicated, and is too lengthy to be discussed here. Details giving the method of constructing the equation for the $2n$-electron system will be found in a paper by Pauling [9] ; Sklar [12] has summarized the method.

Apart from a Coulomb term, subsequently to be allowed for in the S terms, the energy of the lowest state of the π electrons is $-\sqrt{(J_{bc}^2+4J_{ab}^2-2J_{bc}J_{ab})}$. The total energy of the molecule is therefore

$$F=6(C-H)+2S_{ab}+S_{bc}-J_{ab}-\tfrac{1}{2}J_{bc}-\sqrt{(J-_{bc}^2+4J_{ab}^2-2J_{ab}J_{bc})}.$$

Minimizing with respect to x_{ab} and x_{bc} gives

$$x_{ab}=1{\cdot}34,\quad x_{bc}=1{\cdot}44,\quad J_{ab}=50{\cdot}0,\quad J_{bc}=31{\cdot}7,\quad F=987,$$

the units of x being angstroms and of J and F kcal./mol. The energy calculated from the chemical formula and bond energies is 980 kcal./mol. The results are indistinguishable from those obtained by the orbital method.

Bonds of fractional order. Calculations similar to the above may be made for other molecules, but the arithmetic becomes laborious if there are more

than six electrons in the resonance problem. Much labour may be saved, and results little less accurate obtained, by making use of the concept of bonds of fractional order.

Pauling, Brockway and Beach [12] suggested that since the resonance in benzene is mainly between the two Kekulé structures, one of which has double bonds where the other has single, and *vice versa*, the carbon-carbon bonds should be called of order 3/2. They gave a general definition of bond order in terms of the coefficients of the bond structures appearing in the Schrödinger function. Penney [11] has criticized and improved this definition. According to his method, the resonance problem is to be solved by the pair method, and the Schrödinger function Ψ obtained, as if all distances between nearest neighbours were equal. First-order perturbation theory is then to be used to obtain the energy E in the form

$$E=Q+\Sigma b_{ij}J_{ij},$$

where Q is the Coulomb energy and J_{ij} is the exchange integral between the neighbours i and j.

If neighbours i and j were linked by a bond in all structures, the coefficient b_{ij} would be unity; if i and j were not linked by a bond in any structure, b_{ij} would be $-\frac{1}{2}$. Hence the bond order of the resonance problem is defined by linear interpolation as

$$p_{ij}=(1+2b_{ij})/3.$$

To p_{ij} must be added one for the localized bonds.

As an example, consider the resonance problem of benzene. The energy of the lowest state is $(Q+2{\cdot}6\,J)$; since the six electrons are equivalent, $b_{ij}=0{\cdot}43$ and $p_{ij}=0{\cdot}62$. The bond order in benzene is thus 1·62.

Similarly, the bond order in graphite is 1·45. Figure 1 gives a plot of the ionic diameter of carbon as a function of the bond order.

The orders in butadiene work out at 1·91 and 1·33; the lengths, read off figure 1, are 1·34 A. and 1·43 A.; and the resonance energy, read off a similar curve, is 8 kcal.

Coulson [10] has given a definition of bond order in the orbital method very similar to that above. This time the energy of the π-electron resonance problem is expressed

$$E=n\epsilon+\Sigma a_{ij}\beta_{ij}.$$

The energy of a bond is $2\beta_j$, and the energy of no bond at all is 0. Hence the bond order is defined by

$$p_{ij}=2a_{ij}.$$

The energy of the six-electron problem of benzene is $6\epsilon+4\beta$, and therefore $p_{ij}=2/3$. The order in benzene is thus 1·67. The orders in butadiene are 1·89 and 1·45; the lengths 1·34 A. and 1·43 A., and the resonance energy 12 kcal.

A summary of the results on other molecules will be found in the papers quoted above.

Fox and Martin [14] have pointed out that a curve similar to figure 1 may

be drawn connecting the length and the energy of a bond. Reading off the length for naphthalene gives a value 1·41 A , the exact X-ray value.

Hyperconjugation. Only one example is known where the carbon-carbon single bond has a length appreciably different from the normal value 1·54 A. Herzberg, Patat and Verleger [15], from spectroscopic data, find 1·462 A. for the C—C bond in methyl-acetylene $HC{\equiv}C{-}CH_3$. There is no mesomerism of the type described earlier in this section, and no contraction of the single bond would therefore have been expected. The probable reason has been suggested by Mulliken [15]. He points out that from the atomic orbitals of the

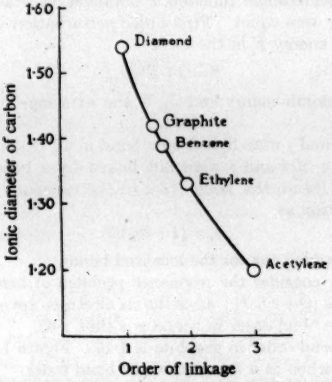


Figure 1. Ionic diameter of carbon as a function of bond order.
(Reproduced by permission of the Council of the Royal Society.)

three H atoms of the methyl group, molecular orbitals may be constructed capable of resonating with the π-electrons of the C atoms. Thus, one set of four resonating orbitals is

$$\psi_a(\pi_+), \quad \psi_b(\pi_+), \quad \psi_c(\pi_+), \quad N(\psi_1+\omega\psi_2+\omega^2\psi_3),$$

where $\omega=e^{2\pi i/3}$, and π_+ stands for a positive angular momentum. Similarly, there is another set of four which are the conjugate complexes of the above. The resonance integral between the last pair of orbitals is so large that the resonance effect is small. Nevertheless, as we have seen in butadiene, even a small amount of resonance produces a considerable shortening of the central single bond. Mulliken calls the effect *hyperconjugation*.

The ethane molecule will show hyperconjugation, but it will be of the second order, and its effect is not sufficient to be the cause of the hindrance to free rotation (see § 4).

Compounds XH. Stehn [17] has made the neat observation that the force constants and internuclear distances in molecules XH (X=B, C, N, O, F) do not change when the molecule is excited. Now an expression for the energy of the state arising from the configuration $(2s)^2\,(2p)^n$ of X may be obtained by the pair method of approximation. The expression involves only Coulomb and exchange integrals for electrons on X, and similar integrals for an electron on X and an electron on H. The former are known from the spectroscopic energy levels of X. The latter are the same in the ground state and the excited states of XH, since the internuclear distances are the same. Hence, the observed energies of the states of XH can be used to determine the integrals. The values of the integrals may be plotted as functions of the atomic number Z, and smooth curves result. The resonance problem is never more difficult than that of four electrons. (This occurs for NH ; the $2s$ shell is filled ; there are three different $2p$ orbits, each containing one electron ; and there is one electron on H.) The integrals so obtained may then be used to predict the positions of other levels.

King [18] has repeated the calculations, using more recent and more accurate spectroscopic data. He gives a complete list of the values of all integrals for molecules XH in the range Li–F inclusive, as well as internuclear distances and force constants in the ground and excited states of the molecules. The energy of dissociation of NH is estimated to be 79 kcal. No direct check on this is possible since NH is not known spectroscopically.

§ 3. THE COLOUR OF ORGANIC COMPOUNDS

An empirical rule states that the intense absorption of light that characterizes dyes will not occur in a molecule which is not mesomeric. However, not all mesomeric molecules are dyes ; for example, the benzene molecule is highly mesomeric, but the great symmetry of its structure leads to the operation of selection rules which prohibit the absorption of visible light.

There are certain groups of atoms whose introduction as units into organic molecules often produces a coloured product. Originally these were all called chromophores ; examples are $-N=N-$, $>C=C<$, $>C=S$, $-N=O$. Recently, however, " chromophore " has tended more and more to mean simply the quinonoid group, present in most dyes.

The presence of a chromophore in an organic molecule is not sufficient to ensure colour. Some such molecules are colourless, or feebly coloured, but strong coloration may be developed by the introduction of certain substitutent groups, called auxochromes. The function of both chromophore and auxochrome is to introduce into the molecule a resonating system of electrons. From this point of view the duality of terms is unnecessary.

Molecules in which the double bonds of a chromophore are conjugated with other double bonds are usually coloured, but the colour does not depend very much on how many double bonds are conjugated. Much more important is the distribution of the bonds, i.e. whether they are arranged in rings, in chains, in rings with side-chains, or in other ways. The colour, in brief, depends on the whole mesomeric pattern.

There are two distinct mechanisms which might be invoked to explain the absorption spectra of dyes, and for a long time it was not possible clearly to decide between them. They are most easily illustrated by comparison with simpler examples of similar types. The first is a direct excitation of an atom localized on a chromophore from one atomic orbital to another. A transition in atomic spectra of this type is $3s\,^2S-3p\,^2P$ in sodium, the well-known D lines. The energy of the transition in the molecule would be affected by valency forces, both in the ground and excited states, and the reason why neighbouring double bonds were necessary would be to keep the energy of the transition low by resonance effects. The second possibility is a change in the coupling of a group of electrons, without any change in the atomic orbitals. An analogous transition, if it were allowed, would be from the ground state to the Heitler-London excited state in the molecule H_2. Here the atomic orbitals in the two states are the same. A further example of this type of transition is that which causes the colour of rare-earth crystals. The transitions are inter-configurational inter-system, and combine various energy states of the incomplete groups $4f^n$ ($n=2, \ldots, 12$).

There has been a general impression, current for some time, that the second of the above possibilities is the proper one. For example, Bury [19] has interpreted in general terms the effect of various auxochromes on certain chromophoric groups, tacitly assuming the second possibility. A satisfactory theory of the colour, or absorption spectrum, of an organic molecule must give detailed values for the energy levels and the intensity of the absorption bands. Only recently has it become clear that calculations by quantum mechanics, based on the second possibility, can give extensive agreement with experiment. The theory has been developed mainly by Sklar [20]. Mulliken [21] has also considered the problem in a series of papers on intensities of electronic transitions in molecular spectra. Sponer, Nordheim, Sklar and Teller [22] have attempted an analysis of the near ultra-violet spectrum of benzene, and give references to most of the earlier work.

The energy levels. Sklar [20] evaluates the energies of various states of benzene, butadiene, fulvene and azulene by methods which follow closely those of Pauling [9]. The value of the exchange integral J is, however, chosen in a different way and some allowance is made for ionic terms.

Following Pauling, the exchange integrals between neighbouring π-orbitals are all taken equal. The resonance energy in benzene is then $2{\cdot}6\,J-1{\cdot}5\,J=1{\cdot}1\,J$, and in butadiene is $\sqrt{3}\,J-1{\cdot}5\,J=0{\cdot}232\,J$. From the following thermochemical data on the energies of hydrogenation of benzene and its products, J may be determined.

$$C_6H_6+3H_2=C_6H_{12}+49{\cdot}8,$$

$$C_6H_8+2H_2=C_6H_{12}+55{\cdot}4,$$

the heat unit being kcal./mol.

A third relation may be deduced from these two, namely,

$$C_6H_6+H_2=C_6H_8-5{\cdot}6.$$

Let B be the hypothetical energy of hydrogenation of a double bond in one Kekulé structure. Then the energy of hydrogenation of benzene to cyclo-hexane is smaller than $3B$ by the resonance energy of benzene. Using the above values

$$1{\cdot}1J+3B=49{\cdot}8.$$

Similarly, the energy of hydrogenation of benzene to cyclo-hexadiene is given by

$$(1{\cdot}1-0{\cdot}232)J+B=-5{\cdot}6.$$

Solving for B and J gives

$$B=32{\cdot}8 \text{ kcal./mol.}, \quad J=-44{\cdot}1 \text{ kcal./mol.}=-1{\cdot}92 \text{ ev.}$$

This value of J agrees well with that of Lloyd and Penney [11] ($-1{\cdot}85$ ev.), but not so well with that of Pauling [9] ($-1{\cdot}55$ ev.).

The energy intervals between the lowest singlet state and the next lowest singlet state in benzene, butadiene, fulvene and azulene are respectively $-2{\cdot}6J$, $-3{\cdot}46J$, $-1{\cdot}76J$ and $-1{\cdot}04J$. The wave-lengths in angstroms of the radiation connected with these transitions, calculated from $J=-1{\cdot}92$ ev., are shown in table 1. Experimental values are given for comparison.

Table 1. *Wave-lengths in angstroms for transitions between singlet states*

Molecule	Calculated	Observed
Benzene	2450	2590
Butadiene	1810	2100
Fulvene	3600	3650
Azulene	6100	7000

The agreement shown in table 1 is remarkable in view of the approximate nature of the theory. Similar agreement is not found for the higher states, especially those that are not singlets. This is indeed only to be expected, since no attempt was made to include the effect of the many excited states. The calculations, however, may be considerably improved by including ionic terms.

Consider benzene as an example. Only the ionic structures of least energy need be included in the Schrödinger function. There are in all twelve such ionic structures, obtained from one or the other Kekulé structure by moving a π-electron in any double bond on to the nucleus to which it was bonded. The secular equation for the energy states now has four unknowns, viz. the Coulomb energy Q and the exchange integral J of the non-ionic structures. the Coulomb energy Q of the ionic structures and an integral K in terms of which the matrix elements of the Hamiltonian operator between any ionic structure and any non-ionic structure may be expressed. The matrix element of unity between an ionic structure and a non-ionic structure is neglected. Since energy differences alone are relevant, only three of the four unknowns matter, and we may take $Q=0$.

The above pair of equations for B and J can be corrected by putting in terms in Q' and K. Two additional equations, which then enable all four unknowns to be found, are obtained from the energy of hydrogenation and the lowest absorption frequency of cyclo-hexene C_6H_{10}. (Actually, the frequency was taken from the spectrum of a similar molecule, since that of cyclo-hexene has not been measured.)

The energy levels of all singlets and triplets within about 10 ev. of the ground state were then calculated. Comparison with experiment showed very fair agreement.

Goeppert-Mayer and Sklar [20] have also calculated the lower excited levels of benzene by constructing the Schrödinger function for the six π-electrons as anti-symmetrized molecular orbitals, i.e. one stage further than the orbital method described in the previous section. They have evaluated all of the important integrals, or taken them from earlier publications, and have estimated the effect of the remainder. The energy levels obtained agree very well with those obtained from the pair method. Details of the calculation of the integrals are given by Sklar and Lyddane [20].

Selection rules and intensities. Group-theory arguments discover very quickly which are the allowed transitions. More difficult to discuss are those transitions forbidden in first approximation, but allowed when coupling of the electronic motion and vibrational motion is considered. Such transitions will, as a rule, be weak. Inter-system, i.e. singlet-triplet, transitions will also occur, but only weakly, and the selection rules for these must be considered.

The general rule for a transition to be allowed is that the matrix element of the electronic dipole moment between the two states shall not vanish. Take axes x, y, z, on each C atom, where x and y lie in the plane of the molecule. Then, in the notation of Wilson [23], the z-component of electronic dipole moment, the nuclei being fixed, transforms like Γ_8, and the x, y components transform like Γ_{12}. If Γ_j and Γ_k are the symmetries of the two states, the transition will be allowed if one of the direct products $\Gamma_j\Gamma_8\Gamma_k$, $\Gamma_j\Gamma_{12}\Gamma_k$ contains the identity representation.

Figure 2 shows the energy diagram of benzene calculated by the pair method, without and with ionic terms. The only allowed transition is ${}^1\Gamma_1$ to ${}^1\Gamma_{12}$, and this is identified with the very strong absorption band at 2000 A. Other bands are also observed, but are very much (several thousand times) weaker. The band at 3400 A. is identified with a transition to ${}^3\Gamma_9$, and the band at 2600 A. with a transition to ${}^1\Gamma_{10}$.

The 2600 A. band is permitted when higher approximations in the theory are considered, because vibrations of the molecule destroy the perfect symmetry. The intensity of the band should increase with temperature, and anomalies in the temperature variation of the electric susceptibility may be expected. Both anticipations are realized. Sponer, Nordheim, Sklar and Teller [22] carry the analysis much further, and even succeed in making a convincing assignment of the vibrational structure in the band.

Mulliken has pointed out in a series of papers [21] that the absorption bands of mesomeric molecules are usually due to a transition between a non-ionic

ground state and an ionic excited state. He constructs molecular orbitals for various states of complicated, but highly symmetrical, mesomeric molecules, in terms of the atomic π-orbitals, and then writes down the matrix element of the dipole moment between different orbitals. Because the π-orbitals are all the same, the expression collapses until only one or two terms are left. Comparison of the expressions for molecules of similar structure, or of one molecule in two configurations such as *cis* and *trans*, gives a direct indication

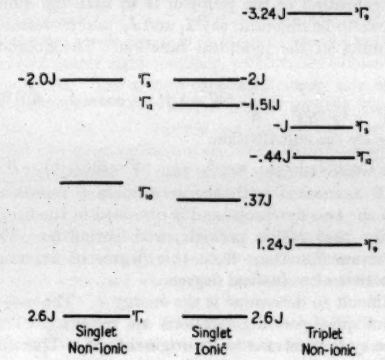


Figure 2. Singlet and triplet energy levels of benzene. (Sklar.)

of relative intensities. Again, intensities of different bands of one molecule may be compared. A molecule as complicated as β-carotene can be handled quite simply, and the results agree well with the observations, as far as they go. Further experiments and improvements in the theory will certainly follow.

§ 4. HINDERED ROTATION AND INVERSION EFFECTS

Hindered rotation

The resistance of single bonds to torsion is usually low, and at ordinary temperatures may even permit free, or nearly free, rotation. This is not the case with multiple bonds. The hindered rotation about a single bond in some molecules is a degree of freedom with an important effect on thermodynamic properties. Several attempts have recently been made to use thermodynamic data to determine the potential function controlling the torsion of single bonds. The ethane molecule H_3C-CH_3 has been most closely studied, because it is simple and the thermodynamic data are now extensive.

Ethane is a symmetrical-top molecule, and the principal moments of inertia do not depend on the relative azimuths ϕ_1, ϕ_2 of the two CH_3 groups about the C—C line. The Schrödinger equation for the rotational motion is in consequence separable. One of the resulting equations refers to the total angular

momentum K, and its component M along an arbitrary axis of quantization. This part of the motion need not be considered in detail, since the associated thermodynamic properties are indistinguishable from those of two classical degrees of freedom. The other equation is the same as the Schrödinger equation for two similar flywheels, coupled by a potential, say $V(\phi_1-\phi_2)$. The function V has trigonal symmetry ; as an approximation, it may be assumed that

$$V=\tfrac{1}{2}A\{1-\cos 3(\phi_1-\phi_2)\}.$$

An obvious generalization of the problem is to take the moments of inertia of the two flywheels to be different, say I_1 and I_2, and to assume ν, instead of 3, maxima and minima in the potential function. The Schrödinger equation is then

$$\frac{1}{I_1}\frac{\partial^2\Psi}{\partial\phi_1^2}+\frac{1}{I_2}\frac{\partial^2\Psi}{\partial\phi_2^2}+\frac{8\pi^2}{h^2}[W-\tfrac{1}{2}A\{1-\cos\nu(\phi_1-\phi_2)\}]\Psi=0.$$

This separates by the substitution

$$\xi=(I_1\phi_1+I_2\phi_2)/(I_1+I_2),\quad \eta=\phi_1-\phi_2,\quad \Psi=\psi(\xi)\chi(\eta),\quad W=E+\epsilon.$$

The energy E associated with the coordinate ξ represents the " solid-body " motion of the two flywheels, and is obtained in the usual way from the boundary condition that $\psi(\xi)$ is periodic, with period 2π. The contribution to the thermodynamic functions from this degree of freedom is also indistinguishable from that of a classical degree.

Much more difficult to determine is the energy ϵ. The η-equation is of the Mathieu type, and quasi-periodic solutions are required. Thus, if ϕ_2 is increased by 2π, the system returns to its original form. Therefore

$$\Psi(\phi_1,\ \phi_2+2\pi)=\Psi(\phi_1,\ \phi_2).$$

But
$$\Psi(\phi_1,\ \phi_2+2\pi)=\psi(\xi+\alpha)\chi(\eta-2\pi),\quad \alpha=2\pi I_2/(I_1+I_2)$$
$$=e^{i\Lambda\alpha}\psi(\xi)\chi(\eta-2\pi),$$

where Λ is the quantum number of the " solid-body " rotation. Hence

$$\chi(\eta-2\pi)=e^{-i\Lambda\alpha}\chi(\eta).$$

This is the boundary condition on the solution $\chi(\eta)$.

The thermodynamic functions of ethane at low temperatures are controlled almost entirely by the six classical degrees of freedom and the internal rotation. When the experimental values are known the effect of the classical degrees is easily subtracted out. Thus to make use of the experimental data, the problem is to calculate the energy levels ϵ of the internal rotation, and hence the thermodynamic properties, in terms of the parameter A. Comparison with experiment then selects the best value of A but, curiously, does not determine the sign. Alternatively, the thermodynamic functions at low temperature may be used to determine the energy levels directly. One constructs the partition function, with the energy levels as unknowns, and chooses them to get the best fit of the data. This method is quite accurate for the first excited level, and possibly the next, provided the statistical weights are known. For ethane, taking unity as the statistical weight of the lowest level,

the statistical weights of the other levels is two, because there are two senses of torsion in all but the ground state.

The quantum mechanics of coupled flywheels was first considered by Nielsen [24]. He showed that the η-equation could be reduced to the form

$$M''+(\alpha+16q\cos 2x)M=0$$

with the boundary condition

$$M(x+2\pi)=(-1)^{A}M(x)$$

in the special case where $I_1=I_2=I$. Perturbation theory gives a good approximation to the energy levels when A is large or small compared with $h^2/8\pi^2I$. The first represents nearly rigid coupling, with small harmonic vibrations about the minimal positions. The second represents nearly free rotation.

Teller and Weigert [25] have evaluated the specific heat of the torsional motion in ethane in the two limiting approximations just described. They have also considered intermediate cases, using tables of eigen-values α of the above Mathieu equation in M and x, for various values of q, computed by Goldstein [26]. These tables only give states of even A, but the roots of odd A lie roughly mid-way between neighbouring even states, and may therefore be obtained approximately without further calculation. Teller and Weigert gave for A a value 360 cal./mol., or 125 cm.$^{-1}$ However, the experimental data at that time were not extensive, and recent calculations give a much greater value for A.

Teller and Topley [27], by considering the equilibrium constant of the reaction $C_2H_4+H_2\rightarrow C_2H_6$, were able to obtain 3000 cal. as a lower limit for A. Howard [28], by considering the selection rules of the infra-red spectrum of ethane, was able to show by comparison with the observed spectrum that free rotation was not possible at ordinary temperatures, and that A was therefore at least several hundred wave-number units.

Pitzer and Kemp [29] have made a much more accurate calculation on the torsional energy levels of ethane, assuming $A\cos 3\phi$ as the potential function, and have prepared tables giving the entropy, heat capacity and specific heat of the internal rotation, for various values of the two parameters A/kT and n^2/IA. The first is the ratio of the height of the potential barrier to kT; the second is effectively that of the energy of the nth state of free rotation of a single flywheel to the height of the barrier. Once again, use was made of eigen-values tabulated by Goldstein. Further eigen-values were calculated for even A, as well as all the necessary ones of odd A. Pitzer and Kemp deduce that A is about 3000 cal. or 1080 cm.$^{-1}$

Kistiakowsky, Locker and Stitt [30] have measured the heat capacities of C_2H_6 and C_2D_6 down to 93° K. Comparing their results with the tables given by Pitzer, they find very good agreement if $A=2750$ cal., or 965 cm.$^{-1}$ The levels of C_2H_6 are 0, 275, 520, 726, . . . cm.$^{-1}$, and of C_2D_6 are 0, 200, . . . cm.$^{-1}$ As a check on the calculations, the heat capacity was evaluated directly from these energy levels, with satisfactory results.

Other examples of restricted rotation. Lennard-Jones and Pike [31] have considered the torsional oscillations in dichlor-ethane CH_2Cl-CH_2Cl, and

similar molecules, with a view to explaining the observed variation of dipole moment with temperature. The Schrödinger equation for the four degrees of rotational motion is not separable, and the moments of inertia depend on $(\phi_1 - \phi_2)$. Beach and Palmer [32] also considered this problem, with results very similar to those of Lennard-Jones and Pike.

Penney and Sutherland [33] have shown that free rotation about the centre single bond is impossible in hydrogen peroxide HO—OH. The angles HOO are about 90°, and the two HO lines are perpendicular. Such a structure seems to agree with X-ray evidence on the liquid [34], and infra-red spectra of the gas [35].

A molecule in which internal rotation almost certainly occurs is dimethylacetylene $H_3C-C\equiv C-CH_3$. Crawford and Rice [36] have measured the specific heat c_p at temperatures 300 to 400 °K., and have compared their results with the tables of Pitzer to find A. They obtain $A < 500$ cal., and the motion is therefore essentially free. Some of the transverse vibrations in this molecule are so low that allowance must be made for them.

Inversion effects and the two-minima problem

The intriguing suggestion was made by Dennison and Uhlenbeck [37] that the nitrogen atom in an ammonia molecule NH_3 is not always on the same side of the H_3 plane. Because there are two positions of the N atom, mirror images

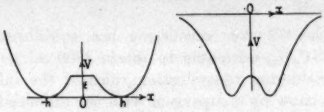


Figure 3. Potential function with two minima. Left, Dennison and Uhlenbeck; right, Manning.

of each other in the H_3 plane, of equal energy, there is a finite probability of the molecule spontaneously inverting itself. The inversion frequency appears in the infra-red and the "wireless" spectrum, and can be measured; it is of the order 2×10^{10} per second, or 0·67 cm^{-1}

A formula can be obtained connecting the inversion frequency with the energy ϵ required to take the N atom through the H_3 plane, and with h, the height of the NH_3 pyramid.

Similarly, a formula can be obtained for the inversion frequency when the molecule is executing the first excited mode of vibration, in which the N atom moves perpendicularly against the H_3 plane. This frequency is greater than the other (26 cm^{-1} as compared with 0·67 cm^{-1}), because the energy of the vibration assists the inversion. From the two formulae and the experimental values of the inversion frequencies, h and ϵ can be obtained; h is about 0·4 A. and ϵ is about 2000 cm^{-1} The inversion frequency for ND_3 may be calculated, and the results obtained agree well with the experimental values.

The mathematical problem is to find the eigen-values in a two-minima

problem. Dennison and Uhlenbeck [37] assumed the potential field shown on the left in figure 3; the two curves are parabolas, whose minima are $2h$ apart. The Schrödinger equation is

$$\frac{d^2\psi}{dx^2}+\frac{8\pi^2\mu}{h^2}[W-V]\psi=0,$$

where μ is approximately the reduced mass of N and 3H. More accurately, μ varies a little with x, but in an unknown way, because the H_3 triangle opens or closes slightly as the N atom passes through. The lowest energy levels are narrow doublets, and an exact formula for the doubling can be found.

Manning [38] has solved the two-minima problem for a potential function V which is more attractive aesthetically than that of Dennison and Uhlenbeck because it has no discontinuities. The form of V is

$$V=-C\,\mathrm{sech}^2\,x/2\rho+D\,\mathrm{sech}^4\,x/2\rho,$$

where C, D, ρ are arbitrary constants. A typical curve is shown on the right in figure 3. The eigen-values W are the roots of a continued fraction, and may be obtained quickly by approximate methods. Manning has used this potential function to estimate the dimensions of NH_3, ND_3 from the observed inversion frequencies. Pitzer [39] has calculated the thermodynamic functions at low temperatures.

§ 5. THE JAHN-TELLER EFFECT

A remarkable theorem discovered by Jahn and Teller [40] states that a non-linear polyatomic molecule cannot at the same time be stable and possess electronic orbital degeneracy. Two applications of this theorem have already been made; one is that the equilateral configuration of H_3 is not stable, and the other is that the octahedron of oxygen atoms surrounding certain paramagnetic ions in crystals is not regular.

The theorem is probably best illustrated by two examples, given by Jahn and Teller.

Linear molecule. The theorem does not in this case preclude the stability of a degenerate orbital state. Consider, for example, the motion of a single electron in the field of three collinear nuclei. The component of angular momentum along the internuclear axis is a constant of the motion. All states except those of zero component momentum are doubly degenerate, because of the two senses of rotation. Suppose that the electron is in a π-state. When a small displacement d of the central nucleus, in a direction perpendicular to the axis, is made, the axial symmetry of the nuclear field is destroyed. The originally degenerate state splits into two. One, of energy E_s, has a Schrödinger function which is even with respect to reflexions in the plane of the nuclei, and the other, of energy E_a, has a Schrödinger function odd for such reflexion. As the displacement d is varied continuously, the functions and their energies vary continuously, but the symmetry properties are preserved. Further, the displacement d is physically equivalent to the displacement $-d$. We must therefore have

$$E_s(d)=E_s(-d),\quad E_a(d)=E_a(-d).$$

The two energies E_s and E_a are even functions of d, and all that is required for stability is that both should have minima at $d=0$.

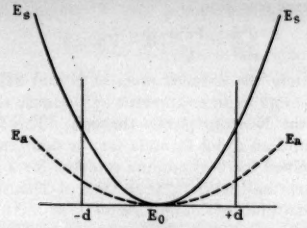


Figure 4.
(Reproduced by permission of the Council of the Royal Society.)

A non-linear molecule. Let ψ be the Schrödinger function for an electron moving in the field of a square configuration of four identical nuclei. Then ψ is a function of the coordinates of the electron and the nuclei. Consider how ψ changes when the coordinates of the nuclei are permuted, corresponding to a rotation of the nuclear square by $\pi/2$ about an axis through the centre of the square and perpendicular to it. Since the physical situation is the same, $|\psi|^2$ is invariant, and ψ has changed merely by a phase factor $e^{i\lambda\pi/2}$. By repeating the process three more times the system returns to its original configuration. Therefore $e^{2\pi i\lambda}=1$, and the values of λ are $0, \pm1, \pm2, \ldots$ The state $\lambda=0$ is single, and all of the others are doubly degenerate.

Let us consider the stability of the degenerate pair $|\lambda|=1$ for displacements of the nuclei shown diagrammatically as I and II.

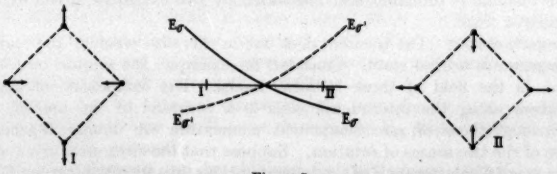


Figure 5.
(Reproduced by permission of the Council of the Royal Society.)

The distortion has in both cases replaced the four-fold axis of symmetry by a two-fold axis. The states have Schrödinger functions ψ_σ, which has a node in the horizontal plane of symmetry, and $\psi_{\sigma'}$, which has a node in the vertical plane of symmetry.

The displacements I and II may be regarded as corresponding to equal but opposite values of one displacement coordinate. For reasons of symmetry

$$E_\sigma(\mathrm{I})=E_{\sigma'}(\mathrm{II}), \quad E_{\sigma'}(\mathrm{II})=E_{\sigma'}(\mathrm{I}).$$

These are the only conditions imposed by symmetry. If E_σ and $E_{\sigma'}$ are expanded as power series in the displacement coordinates, the first-order terms will not in general vanish, because the above conditions do not require the first-order terms to vanish. The energy levels E_σ and $E_{\sigma'}$ will in general cross at the undisturbed position, and this position is therefore not a position of stable equilibrium.

Consider now the non-degenerate state $\lambda=0$. The Schrödinger function has no nodal planes. If E_0 is the energy, then by symmetry

$$E_0(\mathrm{I})=E_0(\mathrm{II})$$

and the energy is an even function of the displacement. For stability E_0 must have a minimum at the symmetrical position.

The general theorem. A brief sketch of the proof of the general theorem is as follows.

Let E_0 be the energy of a degenerate set of levels characterized by the running suffixes ρ, σ. The Schrödinger equation is

$$\mathscr{H}_0\psi_\rho=E_0\psi_\rho. \quad (\rho=1, \ldots n).$$

When the nuclei are given small displacements from the symmetrical configuration the perturbation on the system may be expressed as a power series in the normal coordinates η_r

$$\mathscr{H}_1=\Sigma_r V_r(q)\eta_r+\Sigma_{rs} V_{rs}(q)\eta_r\eta_s+ \ldots,$$

where $V_r(q)$, $V_{rs}(q)$ are functions of the electronic coordinates. We are interested only in the first-order terms since they are the ones that cause instability. The normal coordinates η_r may be chosen in such a way that each V_r transforms according to one or the other of the irreducible representations of the symmetry group of the nuclei in their symmetrical position. Denote these irreducible representations by V.

By first-order perturbation theory the secular equation for the energy levels W, measured from the undisturbed zero E_0, is

$$\mathrm{Det}\,\{H_1(\rho, \sigma)-W\delta(\rho, \sigma)\}=0,$$

where $H_1(\rho, \sigma)$ is the matrix element given by

$$H_1(\rho, \sigma)=\Sigma\eta_r M_{\rho\sigma}, \quad M_{\rho\sigma}=\int\psi_\rho^* V_r\psi_\sigma dt.$$

Clearly, if W is a root of the secular equation for one set of values of η_r, $-W$ is a root when the signs of all η_r are reversed. The symmetrical configuration cannot therefore be stable unless all the roots W are zero. The roots cannot all be zero if one or more of the integrals $M_{\rho\sigma}$ does not vanish.

Let Γ be the irreducible representation subtended by the ψ_ρ. The condition that one or more of the $M_{\rho\sigma}$ does not vanish is that the direct product representations $V[\Gamma^2]$ must contain the identical representation. Thus, a symmetrical configuration with electronic orbital degeneracy is not stable if there exists a normal displacement for which $V[\Gamma^2]$ contains the identical representation. (Totally symmetrical displacements are not to be included, since it may be assumed without loss of generality that the system is stable with

respect to such displacements.) An equivalent statement is that $[\Gamma^2]$ and V must contain the same irreducible representation.

When a molecule is distorted in the direction of a normal mode the perturbation has the same symmetry properties as the vibration. Hence the irreducible representations V are simply those of the normal modes of vibration. Most of these have been given for molecules of varying symmetry in the literature; Jahn and Teller give a complete list, including rotations and translations, for all possible types of molecule (34 point groups in all), and the minimum number of equivalent points, or nuclei, needed to establish the symmetry of each group. A third table gives the reduction $[\Gamma^2]$. A direct comparison of the tables shows that at least one representation occurs both in $[\Gamma^2]$ and the vibrations, and the theorem is proved.

Spin-orbit forces. Jahn [41] has extended the above theorem to cover cases where there is both spin and orbital degeneracy. The magnetic coupling between spin and orbit can stabilize a non-linear configuration which would be unstable if orbital degeneracy alone were present, provided the coupling is very large and the molecular forces split up the lowest state J with a non-degenerate state lowest. Such a situation is exceptional, although a good approximation to it occurs in the Eu^{+++} ion and its six coordinated oxygen atoms in hydrated europium salts.

Jahn also proves that a non-linear symmetrical position of a polyatomic molecule cannot be stable if there is spin degeneracy other than that arising from a single uncoupled spin (Kramer's doubling). This is because the non-diagonal matrix elements of the spin-orbit coupling introduce some of the electronic orbital momentum of the excited states into the ground state, thus causing a very slight instability.

The equilateral H_3 system. The lowest electronic state of H_3 in the equilateral triangle configuration has orbital degeneracy, because the third electron occupies one of the complex ω orbitals (see § 1). The Jahn-Teller theorem therefore comes into operation and tells us that the configuration is not stable.

Paramagnetic crystals. The paramagnetism of hydrated iron-group crystals arises, as is well known [42], from incomplete groups of electrons. A brief account of the theory of the temperature variation of the magnetic susceptibility and related properties of such crystals has been given by Schlapp and Penney in Volume II of these *Reports*.

The main part of the crystalline field acting on the magnetic ion has cubic symmetry, and is caused by a regular, or nearly regular, octahedron of oxygen atoms (water molecules). The forces holding together this complex $X.6H_2O$ are large, as is shown by the observations of Freed [43] and others that the complex still holds together in solution, with its absorption spectrum practically unaltered. As an exploratory approximation, at any rate, it is permissible to treat the ion and its water cluster as an isolated polyatomic molecule.

The perturbing effect of the octahedron of water molecules on the incomplete group of electrons of the paramagnetic ion is purely electrical, and therefore acts only on the charge density of the electrons. Sometimes, but not

often, the perturbation is great enough to orient each electron separately. Usually, however, the exchange forces coupling the electronic motions together are more important than the crystalline forces, and the electrons are oriented as a group. Different arrangements of their total charge density have different energies. Because of the high symmetry of the perturbation it often happens that there are two or three equivalent charge distributions more stable than any others; i.e. the perturbation has not completely "quenched" the orbital angular momentum. The Jahn-Teller theorem in such a case tells us that the six water molecules surrounding the ion are not stable. The octahedron becomes distorted and the orbital degeneracy is removed. The distortion is by no means negligible, since the energy changes caused by small displacements are of the first order in the displacements, instead of the usual second order.

Van Vleck [44] has developed the above ideas mathematically for the three ions Ti^{+++}, V^{+++}, Cr^{+++}, each with its six coordinated water molecules. The first two have a triply degenerate orbital state lowest in the cubic field. The octahedron is distorted so that the X—O distances vary by as much as 0·2 A., and the back-effect of the distortion splits the orbital state by nearly 1000 $cm.^{-1}$ The third ion has a non-degenerate orbital state lowest, and the only degeneracy is that of spin. There will be a very small splitting of the Jahn-Teller type introduced by the orbit-spin coupling in the same way as that considered by Jahn above, but this will be smaller than that caused by the non-cubic part of the field. This latter is estimated at 0·1 $cm.^{-1}$, and is of importance only in demagnetization experiments.

The distortion of the octahedron surrounding the ions Ti^{+++}, V^{+++}, and others with orbitally degenerate states lowest, can occur in two or three ways, differing only in their arrangement in space. Statistically, therefore, the octahedrons will be regular, as shown by X-ray evidence. When circumstances are very favourable there may be a doubling, or trebling, due to "tunnelling", of magnitude comparable with that in NH_3. Such doubling will show up in specific-heat measurements at low temperatures (0–4° K.). Ashmead [45] has in fact observed a peculiar increase with temperature in the specific heat of $CuSO_4.5H_2O$ at about 4° K., but finds nothing similar in $CuSO_4(NH_4)_2SO_4.6H_2O$. Without proper calculations, it is not possible to decide whether this anomaly is caused by tunnelling.

§ 6. MISCELLANEOUS

This section mentions papers which cannot be treated in full, either because of lack of space or because they are not sufficiently closely related to the subject of this *Report*.

Quantal effects in rotational specific heat. The rotational specific heat of ortho- and para-hydrogen at low temperatures ($<60°$ K.) deviates considerably from R, the classical value. An exact quantitative explanation can be given with the aid of quantum mechanics. Other molecules, in principle, show similar effects. Pitzer [39] has considered the most likely of these molecules and has decided that the classical formulae for the rotational specific heat,

the heat capacity and the entropy will give very close agreement with experiment at temperatures above 83° K. (methane), 50° K. (ethylene) and 42° K. (ethane). Of all organic molecules, only methane has sufficient vapour pressure to allow measurements of the departures from the classical values to be made.

Structure of proteins. A summary of recent work on protein molecules will be found in the account of a discussion [48], arranged by the Royal Society. Most of the papers are well off our subject, but one by Wrinch describes, and gives references to, her theory of the cage molecule. Bernal, in the same discussion, and elsewhere, vigorously contests the statement that the only available X-ray data, on the insulin molecule, support the cage structure. Apparently many models would explain the observations better, or equally well. Nevertheless, the cage model is a fascinating idea, and may well lead to a better understanding of the structure of proteins.

Activation energies. Our previous *Report* summarized the method developed by Eyring and collaborators for estimating the activation energies of many reactions. Evans and others [47] have applied and extended the same method to discuss further reactions.

The writer wishes to thank Mr. C. R. Bury for advice on chemical terminology, Drs. Jahn and Teller for permission to use figures 4 and 5 from one of their articles, and the Council of the Royal Society for permission to reprint figures 1, 4 and 5 from the *Proceedings*.

REFERENCES

(1) James and Coolidge. *J. Chem. Phys.* **1**, 825 (1933).
Present. *J. Chem. Phys.* **3**, 122 (1935).
James, Coolidge and Present. *J. Chem. Phys.* **4**, 187 (1936).
Coolidge and James. *J. Chem. Phys.* **6**, 730 (1938).
(2) Pauling and Wilson. *Introduction to Quantum Mechanics.* (McGraw-Hill, 1937.)
(3) Coulson and Duncanson. *Proc. Roy. Soc.* A, **165**, 90 (1938).
(4) Hassé and Baber. *Proc. Camb. Phil. Soc.* **31**, 564 (1935).
(5) Beach. *J. Chem. Phys.* **4**, 353 (1936).
(6) Coulson. *Proc. Camb. Phil. Soc.* **31**, 244 (1935).
(7) Eyring and Rosen. *J. Chem. Phys.* **4**, 121 (1936).
Diamond and Eyring. *J. Chem. Phys.* **5**, 695 (1937).
Stevenson and Hirschfelder. *J. Chem. Phys.* **5**, 933 (1937).
Hirschfelder. *J. Chem. Phys.* **6**, 795 (1938).
Hirschfelder and Weygandt. *J. Chem. Phys.* **6**, 806 (1938).
(8) Hückel. *Phys. Soc. International Conference on Physics, London*, p. 1 (1935). Earlier references are given in this article.
(9) Pauling. *J. Chem. Phys.* **1**, 280 (1933).
Pauling and Wheland. *J. Chem. Phys.* **1**, 362 (1933).
Pauling and Sherman. *J. Chem. Phys.* **1**, 609, 679 (1933).
(10) Lennard-Jones. *Proc. Roy. Soc.* A, **158**, 280 (1937).
Lennard-Jones and Turkevitch. *Proc. Roy. Soc.* A, **158**, 297 (1937).
Wheland. *Proc. Roy. Soc.* A, **164**, 397 (1938).
Coulson. *Proc. Roy. Soc.* A, **164**, 383 (1938).
Coulson. *Proc. Roy. Soc.* A, **169**, 413 (1939).
Lennard-Jones and Coulson. *Trans. Faraday Soc.* **35**, 811 (1939).

(11) PENNEY. *Proc. Roy. Soc.* A, **158**, 318 (1937).
PENNEY and KYNCH. *Proc. Roy. Soc.* A, **164**, 409 (1938).
LLOYD and PENNEY. *Trans. Faraday Soc.* **35**, 835 (1939).
(12) SKLAR. *J. Chem. Phys.* **5**, 669 (1937).
(13) PAULING, BROCKWAY and BEACH. *J. Amer. Chem. Soc.* **57**, 2705 (1935).
(14) FOX and MARTIN. *J. Chem. Soc.* p. 2106 (1938).
(15) HERZBERG, PATAT and VERLEGER. *J. Phys. Chem.* **41**, 123 (1937).
(16) MULLIKEN. *J. Chem. Phys.* **7**, 339 (1939).
(17) STEHN. *J. Chem. Phys.* **5**, 186 (1937).
(18) KING. *J. Chem. Phys.* **6**, 378 (1938).
(19) BUBY. *J. Amer. Chem. Soc.* **57**, 2115 (1935). See also PAULING and GILMAN, *Organic Chemistry*, vol. II, pp. 1888–9. (Wiley and Sons, 1938.)
(20) SKLAR. *J. Chem. Phys.* **5**, 669 (1937).
GOEPPERT-MAYER and SKLAR. *J. Chem. Phys.* **6**, 645 (1938).
SKLAR and LYDDANE. *J. Chem. Phys.* **7**, 374 (1939).
(21) MULLIKEN. *J. Chem. Phys.* **7**, 14, 20, 121, 339, 356, 364 (1939).
(22) SPONER, NORDHEIM, SKLAR and TELLER. *J. Chem. Phys.* **7**, 207 (1939).
(23) WILSON. *Phys. Rev.* **45**, 706 (1934).
(24) NIELSEN. *Phys. Rev.* **40**, 44 (1932).
(25) TELLER and WEIGERT. *Nachr. Ges. Wiss. Göttingen, M.P.Kl.* 218 (1933).
(26) GOLDSTEIN. *Trans. Camb. Phil. Soc.* **23**, 303 (1927).
(27) TELLER and TOPLEY. *J. Chem. Soc.* 885, (1935).
(28) HOWARD. *J. Chem. Phys.* **5**, 451 (1937).
(29) KEMP and PITZER. *J. Chem. Phys.* **4**, 749 (1936); *J. Amer. Chem. Soc.* **59**, 276 (1937).
PITZER. *J. Chem. Phys.* **5**, 471 (1937).
(30) KISTIAKOWSKY, LOCKER and STITT. *J. Chem. Phys.* **7**, 289 (1939).
(31) LENNARD-JONES and PIKE. *Trans. Faraday Soc.* **30**, 830 (1934).
(32) BEACH and PALMER. *J. Chem. Phys.* **6**, 635 (1938).
(33) PENNEY and SUTHERLAND. *Trans. Faraday Soc.* **30**, 898 (1934); *J. Chem. Phys.* **2**, 492 (1934).
(34) RANDALL. *Proc. Roy. Soc.* A. **159**, 83 (1937).
(35) BAILEY and GORDON. *Trans. Faraday Soc.* **34**. 1133 (1938).
(36) CRAWFORD and RICE. *J. Chem. Phys.* **7**, 437 (1939).
(37) DENNISON and UHLENBECK. *Phys. Rev.* **41**, 313 (1933).
(38) MANNING. *J. Chem. Phys.* **3**, 136 (1935).
(39) PITZER. *J. Chem. Phys.* **7**, 251 (1939).
(40) JAHN and TELLER. *Proc. Roy. Soc.* A, **161**, 220 (1937).
(41) JAHN. *Proc. Roy. Soc.* A, **164**, 117 (1938).
(42) VAN VLECK. *Electric and Magnetic Susceptibilities.* (Oxford, 1932.)
(43) FREED and WEISSMAN. *J. Chem. Phys.* **6**, 297 (1938).
(44) VAN VLECK. *J. Chem. Phys.* **7**, 120 (1939).
(45) ASHMEAD. *Nature, Lond.*, **143**. 853 (1939).
(46) Discussion on Proteins. *Proc. Roy. Soc.* A, **170**, 40 (1939).
(47) EVANS and POLANYI. *Trans. Faraday Soc.* **34**, 11 (1938).
EVANS. *Trans. Faraday Soc.* **35**, 824 (1939).

SOUND

By E. G. RICHARDSON, B.A., Ph.D., D.Sc.

§1. FREE PROPAGATION; SUPERSONICS

During 1938 two series of observations on the propagation of sounds in the open air were reported. Kukkamaki [1] carried out measurements of the velocity of sound over tracts of 1 km. in Finland, one series over land and two over a lake, first in summer and then in winter when the lake was frozen. The three sets gave speeds—reduced to 0° C. and absence of humidity or carbon dioxide—which differed by less than 0·1 per cent and had a mean value of 330·79 m./sec. The effect of wind was eliminated in the customary manner, by taking the average of forward and return excursions of the sound pulse, which was produced by the firing of a 1-kg. bomb at 1 m. above the ground. On the other hand the work of Waetzmann,* Scholz and Kruger [2] is concerned with the dissipation in the atmosphere of aeroplane noises. An aeroplane passed at various heights in horizontal flight over a microphone so that the sound intensity could be recorded when it was exactly overhead. From these data an intensity : height curve was plotted. Atmospheric losses were indicated by departures from the parabola which an inverse square law would give. These were attributed to the water vapour and carbon dioxide in the atmosphere. Apart from increased decrement, the curve for a humid day did not indicate any abnormality in propagation.

Colwell, Friend and McGraw [3] describe an apparatus which should be of use in a senior school or university laboratory. A glow-discharge valve working in a simple circuit off the A.C. mains (60 c./sec.) causes a series of short pulses at the same frequency to be emitted from a loud-speaker. At a short distance along the bench is a microphone from which the current is led to the vertical pair of plates of a cathode-ray oscillograph while the sweep circuit of the instrument sweeps the electron spot from left to right at the same rate of 60 times a second. The sound pulse when picked up appears as a vertical jag in the horizontal line traced by the spot on the screen, and the position of this jag is governed by the relative epochs of emission and arrival of the pulses. The distance that the microphone must be moved between two stations that give the jag in the centre of the oscilloscope screen is equal to the wave-length of a 60 c./sec. tone in the air of the laboratory, so that the experiment takes up five or six metres of bench space, but this can be reduced, at the loss of a little accuracy, if the emitting circuit is arranged to give harmonic pulses of the driving frequency.

* The science of acoustics has suffered a grievous loss by the deaths, within a few months of each other, of two professors who devoted their lives to research in this subject, E. Waetzmann of Breslau and Max Wien of Jena; and of a younger physicist of great promise, J. E. R. Constable of the National Physical Laboratory.

A theoretical paper by Herzfeld [4] considers the modification introduced into a wave reflected in a fluid by the conditions at the solid boundary which produces the reflection. The factors which he considers are the scattering of impinging molecules by roughness of the surface, adsorption and thermal conduction of the solid. The latter is the most important since its effect is to give rise to another reflected wave besides the regularly reflected one. While the extraordinary wave dies out in a small distance from the surface, the ordinary reflected wave has its overall amplitude reduced in virtue of the fact that both reflected amplitudes must add up to the incident one at an unyielding boundary. The result of this should be that the fluid molecules near the reflector should have a higher temperature than those in the body of the fluid. Experiments are in hand to verify this. With a good conductor like brass or copper as reflector, the loss in reflected amplitude may be considerable, and avoidance of such reflectors when making precise measurements of stationary waves in fluids is indicated. With insulators like wood and ebonite the effect is not appreciable. Another point which might influence supersonic measurement in gases has been considered by Santon [5], who gives a form of the Herzfeld-Rice relaxation equations modified to allow for viscosity, though he doubts whether the difference would fall outside the limits of experimental error.

In a sequence of six papers, van Itterbeek, van Paemel and Thys [6] report measurements of absorption and its variation with pressure in a number of gases. The present authors have worked on carbon dioxide and nitrous oxide, as well as on oxygen, hydrogen, deuterium and argon at temperatures near their critical points. In the latter group the absorption coefficient falls with rising pressure, in agreement with classical theory, but—at least in their range of frequency and pressure—the reverse is found for carbon dioxide. The absolute value of the absorption is, as one now expects, greater than the Stokes-Kirchhoff value—twenty times in hydrogen, ten times in deuterium, four times in helium, three times in neon at ordinary temperatures—but all the values decrease at low temperatures. When a magnetic field of 6 k gauss is applied at right angles to the supersonic field in oxygen, the absorption decreases by 20 per cent at N.T.P., though the velocity is apparently unchanged. It should be pointed out that since these absorption coefficients were measured by direct reaction of the reflected waves on the oscillator—a method usually reserved for wave-length determination—they must be regarded as approximate. The measurements of Schmidtmuller [7] on the one hand and of Knudsen and Fricke [8] on the other are free from this objection, since in the former case a pick-up was used to measure intensities in a direct line from the source and in the latter the average intensity in a miniature reverberation chamber was recorded on a swinging microphone. Besides pure nitrogen, oxygen, carbon dioxide and carbon disulphide, these authors have employed mixtures of these gases. Other work on mixtures has been done by Kuchler [9] and Metter [10].

For velocities Jatkar [11] has begun a series of researches on the best form that the supersonic interferometer should take. He brings out an important caveat for the user of such apparatus. It is common to let the quartz oscillator radiate through a hole in one of its electrodes. If the space between it and the

reflector is small, i.e., if a high-frequency source is radiating into a rather narrow tube, the system behaves like a tiny Helmholtz resonator. The maxima of reaction on the quartz occur, not when the separation of the quartz and the reflector is an integral number of half wave-lengths, but when the *volume* of the intervening space passes through critical values. He also shows how reflections from the walls of the interferometer tube modulate the form of the "peaks". Altogether these papers form a valuable *vade mecum* to anyone engaging in supersonic velocity measurements by the Pierce method. The effect of containing tubes on supersonic velocities (this time for liquids) is discussed by Hulswit [12] in reference to radial resonance. "If ultrasonic measurements are to be relied upon to give the unconfined velocity, the dimensions of the vessel must be very large in comparison with the wave-length." Of course it is possible to reduce radial resonance by lining the tube with absorbent material, as Jatkar did on occasion.

For moderate-sized tubes Bürk and Lichte [13] have shown that the friction on the walls causes a modification of plane waves to an extent which depends on the frequency and so produces a marked distortion in a complex sound sent through the tube. They used an iron pipe 14 cm. wide and 7 m. long and picked up the sound from a loud-speaker on microphones at various stations along the pipe. Irregularities first appeared when the sender frequency exceeded 3000 c./sec. As the frequency rose up to this critical value the ratio of output to input remained practically unchanged, but when this frequency was exceeded there were wide fluctuations in the efficiency of the transmitting system from one frequency to another. Furthermore, the velocity of these high-pitched sounds was below normal. The phenomenon is also ascribed to radial waves. At the other extreme of very narrow tubes there is, of course, low velocity and high absorption from viscous drag on the confines of the tube. A question of some importance for high-frequency propagation in narrow channels has been answered by May [14], who, in a series of measurements on capillary tubes with sources of monel needles in magneto-strictive oscillation, has shown that though the wall effect is greater than that to be expected from classical theory, it is not abnormal in comparison with low-frequency absorption in such tubes. In each case a viscosity of about ten times the static value substituted in the usual Kirchhoff expression agrees with the experimental observations. Continuing his work on the resonance frequencies in liquids in (wide) tubes in collaboration with Istraty, Salceanu [15] finds a large number of secondary resonant lengths besides the principal ones predicted from the known velocity of sound in the liquids. Possibly these too are radial vibrations of the liquid in sympathy with overtones of the source.

Turning now to *supersonic* propagation in liquids, and passing over those papers which deal with the *theory* of light diffraction by supersonics, we find that the practice of this useful method of measuring supersonic velocity does not languish. A number of workers are concerned in finding dispersion at very high (hypersonic) frequencies with diverse results. This is evidently due to the difficulty of making observations with frequencies of billions of cycles per second, where the diffracting grating of wave-fronts has a spacing of a fraction

of a millimetre. In the circumstances we can do no more than catalogue the results, in the hope that more precision may eventually be brought to bear on them.

Author	Liquid	Frequency range (Megacycles/sec)	Velocity change (per cent)
Dutta [16]	Water	1 to 30	0·2
,,	Xylol	10 to 30	none
Spakowsky [17]	Water	0·2 to 3	0·5*
,,	Acetic acid	,,	1
Rao [18]	Carbon tetrachloride	7 to 5000	10
,,	Acetone	,,	22
Bär [19]	Water	7 to 53	none
,,	Ethyl acetate	,,	none
,,	Benzene	,,	none

* From the graphed results; wrongly quoted in the paper as 5 per cent.

There are also velocity determinations by Liepmann [20] in liquid oxygen, Hirschlaff [21] in liquid nitrogen, Yosioka [22] in heavy water, Findlay [23] and co-workers in helium and Biquard [24] for a number of organic liquids under pressure (velocity varies linearly with pressure).

These are all with the optical diffraction method, which has almost become standard for velocity measurements. Seifen [25] describes a precision form of the apparatus, in which he attempts to improve the accuracy of 0·1 per cent previously claimed for it. The principal points to which he looks for betterment are (*a*) stabilisation and exact measurement of frequency of supersonic source. (*b*) temperature constancy of liquid cell, (*c*) standardisation of screw which moves cell across light field. As regards (*a*), one of many harmonics of a 100 kc./sec. source linked with a quartz detector *in vacuo* is picked out. The quartz controller is compared from time to time with the wireless signal from Brussels upon which broadcasting frequencies are based. For (*b*) a thermostat, and for (*c*) comparison with a standard metre served. Velocities in this apparatus can be estimated to an accuracy of 1 in 6×10^5 at a frequency of 5 megacycles per sec.

Bergmann and Oertel [26] project the shadow of a supersonic diffraction grating, formed behind a trough of liquid in which progressive waves are set up, upon a rotating mirror before it reaches the usual matt screen. The diffraction pattern is moving, let us say, from left to right of the screen at a high speed when the mirror is still, but by rotating the mirror from right to left at an appropriate speed the pattern may be "held" on the screen. The supersonic velocity in the liquid is then calculated from the speed of rotation of the mirror.

A number of authors, e.g., Grobe [27], Bazulin [28], attempt to measure the decay of amplitude of a supersonic beam in a liquid by comparing the light diffraction patterns at various distances from the source, on the assumption

that the intensity of the first diffracted image is proportional to the supersonic amplitude at the station in question. In view of the difficulties in measuring the optical intensities, and the complexities in the theory of the effect, which may invalidate the assumption, it is not surprising that widely differing values for the absorption coefficients are quoted. Nor are the mechanical methods in which amplitude is measured by the pressure on a pendulum in the liquid in much better case, because there is often an additional pressure due to a current of fluid going out from the centre ("electric wind"), and the presence of a pendulum bob is found to give rise to gratuitous reflections. Probably Buss [29] has brought the technique of the radiation pressure method to its highest point yet in his work on the absorption in a stationary wave system, since he has devised a sensitive valve circuit which embodies the pendulum as one side of a condenser. It is amusing to note that while Grobe has no faith in the pendulum method, Buss condemns the optical method because he can find no connection between supersonic amplitude and diffracted light intensity. We are inclined to agree with both of them!

Cance [30] has a novel method for velocity which depends on the ripples set up on the free surface of a liquid by two quartz emitters so immersed as to make an obtuse angle, 2α, with each other. The displacement at the surface depends on the interference of the two components of pressure at this almost complete reflector, and raises a system of ripples which is the acoustical analogue of Lloyd's mirror fringes. The spacing of the ripples is therefore equal to $\lambda/2 \cos \alpha$. If the fringes are traversed with a microscope, the wave-length, λ, of the radiation within the liquid can be calculated. Another phenomenon which can be readily demonstrated by supersonic ripples is that of the reflected and refracted waves set up when radiation strikes the boundary between two media of different elastic properties. Kretschmer and Rschevkin [31] send supersonic ripples across such a boundary and are able to demonstrate the existence of all three sets of waves by stroboscopic illumination; also total reflection under suitable conditions.

Several papers in 1938 deal with polarised light in transparent substances carrying supersonic radiation. Lucas [32] has found that liquids so irradiated show the property of double refraction, which he ascribes to the anisotropy set up by the supersonic waves, a phenomenon closely linked with the discovery of Maxwell that a velocity gradient in a fluid renders it doubly refracting. With crossed nicols and a beam of light parallel to the emitting face of the quartz, one gets a restoration of light when the quartz is set in oscillation. The effect is a maximum when the principal planes of section of the polariser and analyser are at 45 degrees to the supersonic wave-front. Resin and linseed oils show it well. Hiedemann and his co-workers had already demonstrated the effect in glass (cf. *Reports*, **3**, 109), and Hoesch [33] now shows how it can be adapted to determine the elastic properties of glass and their variation with temperature.

The interaction of supersonics and colloids continues to intrigue several workers. Laidler and Richardson [34] have estimated the absorption suffered by radiation at three frequencies between 40 and 700 kc./sec. in smokes of magnesium oxide and stearic acid formed naturally by combustion and in artificial

aerosols of uniform particle size, made of the spores of several fungi. Checks are made with the theories of Rayleigh and Sewell for the absorption of sounds in fogs. Whereas the relationship is verified in respect of size and concentration of the disperse phase, the theory does not account for the observation that the attenuation increases with frequency. Two investigations on the effects of supersonics on colloidal viscosity lead to rather different results. Freundlich and Gillings [35] have established two important facts. Firstly, it is the cavitation induced in liquids by the high-frequency agitation which is the operative factor in lowering viscosity after exposure, so that if the formation of vacuous cavities is inhibited by the application of external pressure little change in fluidity takes place. The second point is that those hydrosols whose viscosity can be lowered are all ones which exhibit anomalous viscosity, i.e., viscosity varying with rate of shear. Since these sols have a noticeable structure, like rubber latex and common jelly, the role of cavitation in promoting rupture of the structural links is not difficult to conceive. Ordinary liquids suffer no such change, provided the heat arising from the absorption of the supersonic energy is dissipated in some way without raising the temperature of the liquid. Thieme [36], on the other hand, who has also studied gelatine and agar sols, presents a less harmonious picture. The viscosity does not decrease exponentially with time, as one would expect if the action were a regular breakdown of gel structure, but its fall is rather incoherent. Moreover the initial water content has a mysterious influence on the trend of events, there being a most favourable concentration for a swift dissolution of the links. It must be borne in mind that with the concentric cylinder viscometer commonly favoured by workers in this field a true value of the viscosity is difficult to arrive at, since the rate of shear on the liquid varies across the interspace at every instant. This may account for some of the curious results obtained by Thieme. To cavitation also Söllner [37] ascribes the action of supersonics in bringing a solid into suspension in a finely divided state in a liquid, a process of increasing technical importance.

Hermans [38] elaborates the theory of a new effect first suggested by Debye. When colloidal particles are charged, the double layer surrounding them should be deformed in a sound field, as they lag behind the motion of the solvent and become flattened like an oil-drop under centrifugal force. A dipole moment is thus set up whose magnitude depends on the local sonic amplitude. In stationary waves there will accordingly be an electrostatic potential difference between neighbouring nodes and antinodes. This difference will be heightened if the radius of an ion is large compared with the dimensions of the double layer, since the flow round the particle may then, in Hermans' opinion, deviate from streamline flow.

We must pass over certain papers—becoming fewer of late—on technical applications of supersonics, dealing with problems of more interest to the industrial chemist than the physicist, but we should draw attention to some uses of magneto-strictive oscillators in the testing of materials. For instance, having two such oscillators clamped at either end of a concrete beam, one to act as sender and the other as receiver, Meyer and Buchmann [39] were able to

detect internal cracks in the beam by the small amount of energy that passes such fractures. If a cement mix is allowed to solidify between the two oscillators, the rate of solidification is paralleled by the rise in the transmitted energy with time.

Fox and Rock [40] have tackled a problem of some importance in supersonic interferometry by exploring the sound field in front of a quartz disc radiating into a liquid, though again it is the pressure on a small glass bead suspended by a quartz fibre which is used. Their apparatus is shown in figure 1, in which most of the letters are explained. The essentials are the quartz oscillator, H (2500 kc./sec.), the little sphere, L, capable of being traversed across and through the supersonic field set up in the liquid in the tank, and the diffusing

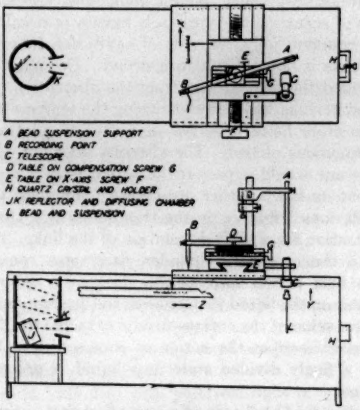

A BEAD SUSPENSION SUPPORT
B RECORDING POINT
C TELESCOPE
D TABLE ON COMPENSATION SCREW G
E TABLE ON X-AXIS SCREW F
H QUARTZ CRYSTAL AND HOLDER
J K REFLECTOR AND DIFFUSING CHAMBER
L BEAD AND SUSPENSION

Figure 1.

chamber on the left to catch the radiation not absorbed in the liquid and to carry it out of the field by repeated reflection from wall to wall of the vertical chamber. These walls were coated inside with coarse gravel to assist the dissipation of the entrapped energy. This device is thus the acoustical analogue of the black-body absorber of optical experiments. Before its introduction standing waves spoilt the observations. In spite of the difficulty of the measurements and the occasional erratic behaviour of the detector, general agreement with the theoretical pressure distribution for the plane piston radiator was obtained. Other difficult diffraction problems concerning obstacles in a sound field have been solved theoretically by Muller, Black and Davis [41] of the Bell Telephone Laboratories. Rayleigh gave the velocity potential for a *spherical* obstacle, and they have now provided solutions for the cylinder and the cube.

Bate [42] has revived an old subject of enquiry, the whispering gallery. He

carried out some experiments in the gallery beneath the dome of St. Paul's Cathedral, London, using heavy tuning forks capable of vibrating for some time and ranging in frequency from 256 to 6000 c./sec. They were held in turn at a fixed distance from the wall and in every case the sound was heard by listeners near the wall in others parts of the gallery. The author thinks that the facility with which a whisper travels without loss is not due to its high-frequency components but to the small amplitude which limits the range within which the direct sound can be heard, so that this does not mask that which goes round the wall. The sound heard, then, is that which has travelled round the wall by successive reflections along short chords. Incident and reflected waves combine to produce wave-fronts travelling at right angles to the wall. At the wall they are in phase but a short distance out from it they are out of phase and come into phase again further towards the centre of the dome. The general conditions necessary for a good whispering gallery are (1) a hard wall for good reflections, (2) a source of low intensity, (3) a source near the wall and directed so that the incident wave falls on the wall at nearly grazing incidence.

§ 2. VIBRATING SYSTEMS: SOLIDS

On the properties of vibrating systems in general we find an interesting account by Waetzmann and Kurtz [43] of their study of undertones, forced vibrations in which the response to the system contains subharmonics of the exciting frequency. They can be produced in any solid body, be it wire, bar or plate, in transverse vibration by the use of special means of support or excitation. Familiar cases occur in Melde's experiment or when the stem of a tuning fork is pressed on the table. They have been able as well to demonstrate subharmonics in the carbon microphone, where it is the pressure of the granules on the diaphragm which is the special condition for their existence, and in an ordinary telephone diaphragm if the exciting magnets are brought near enough to touch the membrane. With the exception of Melde's experiment the common characteristic is a one-sided support or constraint of the body in forced vibration. The fundamental equation of motion must contain parameters (mass, damping or stiffness) of which the values are functions of the displacement, instead of being regarded as constant, as they are in the elementary theory of forced vibrations.

With regard to standards of frequency, there are two papers which deserve our attention. Schuck [44] has constructed a valve-maintained tuning fork with loads whose position can be adjusted without stopping the fork. The weights are pushed or pulled along the prongs by springs, controlled from a pointer through a link motion in such a fashion that a linear scale of frequency is obtained on the dial over which the pointer moves. This has the effect that correction for temperature merely involves a bodily movement of the scale, since over a small range of temperature the change of frequency is in direct proportionality. This shift is effected by turning a screw or automatically by a bi-metallic strip. The accuracy is 0·06 per cent. The frequency and time standard of Essen [45] is a quartz resonator in the form of a ring whose axis

coincides with the optic axis of the crystal and has a very small decrement (10^{-5}). The electrodes are metal rings, one inside and one outside the quartz ring. The electrodes each carry on an extension three phosphor bronze screws which locate in grooves cut radially in the quartz at three nodal points, and these remain fixed in position with respect to the quartz. The crystal is self-maintained in a Pierce circuit and its stability assured by making it drive a clock whose rate of working can be predicted to within 0·001 sec. a day.

This paper leads to a consideration of a number of others on the construction of high-frequency oscillators and the velocity of waves of compression in solids. Ni Tsi Ze and Chung Ming-Sang [46] find a number of interesting modes of vibration in the hollow quartz cylinder, including those in which it acts like a bell (combined radial and circumferential nodes), of which they reproduce some Chladni figures. Zacek and Ptrzilka [47] have also treated the same problem both in theory and experiment, including torsional as well as radial vibrations. In the former case the fundamental mode has a higher pitch than that of a complete disc when the ring becomes sufficiently narrow, but the reverse holds for the radial oscillations. The nodal lines of both are circles. They have shown that the little air column within the hollow cylinder can resonate, if of suitable dimensions. The existence of these aerial vibrations can be shown by strewing powder within the quartz.

Pierce and Noyes [48] also describe a number of circuits useful for maintaining h.f. sources and for detecting weak vibrations. The circuits include one of a simple and easily fitted-up type for maintaining rods in magneto-strictive oscillation, to which we advise those who contemplate using such sources to give their best attention. Focke, Lindsay and Wilks [49] have measured velocities in single crystals of bismuth (about 10 cm. long) both where the principal axis of the crystal was perpendicular to the direction of propagation and in others where they were parallel. After the crystal has been grown it is inserted between two steel sections. One of these steel rods is supplied with radiation through a loud-speaker and 15-watt amplifier, while each has near it a piezo-electric gramophone pick-up. Hence both frequency and the position of the nodes are determined. A simple relation enables the acoustic impedances of the three component solids to be found and hence the velocity in the central one. The method is thought to be particularly adapted to determining acoustic wave velocities in small specimens of brittle substances which cannot stand much strain. Baumgardt [50] proposes similar measurements with the crystal between two liquid strata and with supersonics of such frequency as to excite resonance in the crystal. This condition corresponds to a maximum in the transmitted energy. He has already applied the idea successfully to steel plates. The method was, as a matter of fact, proposed by Bär and Walti [51] several years ago, and the second collaborator now describes his arrangement for the apparatus. The thickness of the solid (glass) through which the supersonic beam has to pass is varied by making it in the form of a wedge and pushing it into the path of the supersonic beam until a maximum or minimum in the transmitted energy occurs. To make this test one observes the extent of the light diffraction by the liquid behind the plate. The increase of thickness corresponding to a change

from one maximum to the next is then equal to a half wave-length of the supersonics in the material. The plate can also be set for oblique incidence of the supersonics, and Reissner [52] and Bär [53] have evolved elaborate theories to explain the observations in such a case; so have Levi and Nath [54], who take account of multiple reflections. The mathematics is too long to abstract here, but a special case of the problem had already been treated by Rayleigh in his *Sound*, viz., the oblique incidence of sound upon a series of plane parallel layers of fluid.

Pierce and Noyes [55] have measured the fundamental frequencies of solid steel cones in longitudinal vibration, comparing with those of uniform cylindrical rods, and using magneto-striction. The theoretical values are given by the roots of the equation

$$\tan \omega l/c = \omega l/c,$$

as for a conical column of air (ω=pulsatance, l=length, c=velocity), and agree quite well with experiment. They have also derived the corresponding formula for a truncated cone. For a given ratio of end diameters the overtone frequencies are not integral multiples of the fundamental, but the difference for successive overtones ($\omega_{n+1}-\omega_n$) is nearly a constant quantity as n is put equal to 1, 2, 3, 4 etc. Thus in one case

$$\omega_n/\omega_1 = 1,\ 1{\cdot}719,\ 2{\cdot}427,\ 3{\cdot}310,\ 3{\cdot}833,\ 4{\cdot}534$$

and

$$(\omega_{n+1}-\omega_n)/\omega_1 = \quad 0{\cdot}719,\ 0{\cdot}708,\ 0{\cdot}703,\ 0{\cdot}703,\ 0{\cdot}701.$$

The remaining papers on vibrating solids concern audible frequencies. Allen [56], noticing that the values for the velocity of sound in copper quoted in tables of physical constants are sadly out of date and refer presumably to material of less purity than one can obtain today, has determined the natural frequencies of work-hardened copper rods. The velocity deduced falls from 3852 m./sec. at 0° c. to 3815 m./sec. at 100° c., giving a coefficient of -9×10^{-5}. The velocity derived from Young's modulus and density of the material is 3852 m./sec. at room temperature. Landon and Davies [57] have experimented in the same field, their object being to determine the modulus of rigidity for nickel and its alloys at room temperature. Torsional vibrations were imposed on a bar of the metal by alternating magnetic fields. The appropriate expression connecting the resonant pulsatance with dimensions, rigidity and density is derived and a size chosen to give a resonance in the neighbourhood of 1000 c./sec. Loads were added to cut down the amplitude to something of the order of 10^{-3} radian, to avoid stressing the bar beyond its elastic limit. The theoretical accuracy in the final result is one in five thousand. Actually the discrepancy between individual results is less than one-fifth per cent. These new data, too, will be of value to engineers, since only meagre and inconsistent results for pure nickel previously existed.

The old controversy on what constitutes a good violin still continues. A report by Jarnak [58] on improvements in the resonating qualities of violins is of special interest because the author is himself an instrument maker and during the tenure of an Oersted Fellowship was able to demonstrate the

improvements he had effected by making and analysing oscillograph records at the Bell Telephone Laboratories. He divides the body of the violin into six resonating parts or sound-boards and adjusts the relative positions of the bridge and the sound-post, as well as giving attention to the shape and thickness of the wooden boards, to get the best distribution of resonances between the lowest note of the instrument and an upper limit of 2000 m./sec. The criterion for a good amplifier is that its resonant peaks should be evenly distributed throughout the gamut and that it should give prominence to the fundamental in any note which it is called upon to radiate. Poor violins have an excess of overtones, especially of the fifth partial, overshadowing the fundamental. Thus, by moving the bridge of a 'cello so as to decrease the tension on the three highest-tuned wooden parts the amplitude of two peaks between 1000 and 2000 c./sec. was lowered; while after bridge and sound-post had been set for a quality most pleasant to the ear, a new peak appeared in the neighbourhood of 1000 c./sec., all peaks then being of a size. The author hopes to be able to reintroduce the obsolete five-stringed 'cello, for which J. S. Bac wrote, and which facilitates the execution and allows of a strongly conected nasand well-tuned resonating body. The other violin papers are by Meinel [59], whose work we have previously reported. He is more concerned with balance of intensity in passing from one note to another. On a poor violin there is great variation in intensity, in spite of even bowing, but on a good old Italian model the loudness curves, though still showing ups and downs, preserve a general level of contour. In his opinion it is only the tolerance of the ear and the fact that the instrument gives out complex and not simple tones that saves this unevenness from being apprehended.

For the struck string there are only theoretical papers by M. Ghosh [60] to notice. He has extended the well-known Kaufmann theory and arrived at an expression for the displacement of a point on the string which is independent of certain assumptions in the original theory.

Both Mary Waller [61] and Colwell [62] and his collaborators continue their work on Chladni figures. The latter show that the figures of several complicated polygonal plates can be built up from those of the component triangles into which the configuration can be divided. The former has been testing Chladni's Law for circular plates, i.e. that the addition of two nodal diameters raises the pitch by approximately the same amount as the addition of one nodal circle. Actually this addition is not a constant quantity; from two to five nodal diameters are required to equal one circle as the complexity of the pattern increases. Abnormal patterns are due to the combination of two nearby normal nodes. Comparison has also been made with Chladni's original drawings in *Die Akustik* (1802). Miss Waller has also carried out preliminary experiments with cymbals, which will be continued when a powder has been found which does not slide off the slanting plates.

It would be odd if a year passed without a reference to the mysterious "strike note" of bells. This year it is in a very readable contribution from Holland by Jan Arts [63], entitled "Jottings from my experiences in the domain of the sound of bells". He opines that the fifth partial suggests the

strike note to the ear. He sometimes observes another subjective tone " which has the peculiarity (*a*) it is very modest in its appearance (*b*) it does not react upon a tuning fork. It seems to be related to the third partial, since it has the same decay time ".

Finally, K. N. Rao [64] has continued Raman's work on the Indian drum (*mridanga*) whose membrane is loaded with iron filings and, on the assumption that the loading varies inversely as some power of the radius, shows how the modes may become harmonic.

§ 3. VIBRATING SYSTEMS: FLUIDS

As usual there is quite a number of papers on tones initiated by the wind. Holle [65] has made a painstaking research of aeolian tones, the sounds produced when a wire is exposed to a draught, paying particular attention to the quality and intensity of the sound. The latter is, as we should expect, a linear function of the wind velocity, while spectrum analyses of the note given out show it to be remarkably free from overtones. Coop [66], on the other hand, places the wire in an *ignited* gas jet and—in contradistinction to Carrière, who has already (unbeknown to this author) investigated the identical phenomenon—decides that these tones are likewise formed by the vortices *behind* the wire. The frequency should then be a function of the diameter of the wire and the velocity of the gas jet, though it does in fact vary also with the distance of the wire from the burner orifice, as if it were an edge tone. On the other hand, the approach of a genuine " edge " in the shape of a metal plate behind the wire stops the sound. The wire itself did not vibrate and the tone was heard only while the gas was burning. When a wire heated bright red by an electric current was placed in the unignited gas jet a very faint sound could be heard. It was believed that the oscillations were present in the unignited gas and that the tone in an ignited jet is due to miniature explosions as the vortices are formed behind the wire ; but the phenomenon is still somewhat obscure. Suzuki and Miduno [67] have been struck by the considerable sound emanating from the thin ribbon of cane or whalebone which is often stretched by Japanese across the back of a flying kite. They were attracted to this question not only by natural scientific curiosity but with hopes of applying the results of their researches to the measurement of the velocity of the wind at places well above the ground. For the actual tests they employed magnesium ribbons set broadside on to the draught of a wind tunnel. Besides transverse vibrations of the ribbon in the direction of the wind, torsional modes were also observed and more frequently than the former, evidently because of the usual formation of vortices alternately at each edge of the ribbon. Again, the note was nearly all fundamental, and its intensity rose nearly in proportion with the wind velocity. The frequency, however, was always that of the fundamental of the ribbon and—unlike the aeolian tones of a wire—independent of velocity over a wide range.

G. B. Brown [68], who writes in a style very refreshing to one who reads through many scientific papers, gives under the title " Jets musically inclined " a résumé of wind-tone theory and practice ranging from Leonardo da Vinci to his own extensive researches in this subject. The paper would be spoilt by

attempts to abstract it, so we must be content to recommend it to those interested in the behaviour of jets and the *modus operandi* of organ pipes and who desire to see some ordering of what may still be described, in the words of the late Lord Rayleigh, as "a vast accumulation of isolated facts and measurements which lie as a sort of dead weight on the scientific stomach". G. B. Brown has done not a little himself "to mitigate this untoward dyspepsia".

Passing now to air columns, there is quite a large group of works on the acoustics of wind instruments of various kinds. Spectrographs of organ pipes are reproduced by the Trendelenburg-Thienhaus-Franz [69] combination, who have been working in this line for several years, as well as by Bonar [70]. The former have now turned to the vox humana, a stop of organ pipes each note of which is a pair slightly out of tune to produce a quavering effect, which at its best imitates a human singer and at its worst reminds one of the goat! The beat frequency comes out well in the spectrum. They have also recorded the best stops of a number of famous German organs. The marked differences in the initial timbre before the pipe comes fully "on speech" are apparent not only in playing isolated notes but in the oscillograms of short set pieces (of Bach and Buxtehude). The influence of the acoustics of the church on the records is discussed, but it is a pity that the authors do not deal more in scientific fundamentals. It should surely be possible to correlate the wave-forms with the shape of the air column and the theory of edge tones as it applies to the flue pipe. The same criticism applies to Bonar's work on reed organ pipes, which is almost entirely a matter of collecting data.

Of more basic outlook are the papers of Mokhtar [71] on flue pipes, Auger [72] on reed pipes, and R. N. Ghosh [73] on the clarinet. Mokhtar's oscillograms of the sound from an open diapason pipe at various blowing pressures exhibit the relationship between the emitted sound and the edge tones at the mouth. At normal pressure the fundamental predominates but the octave may be elicited either by increasing or reducing the pressure from normal. Auger used a cone 1 m. long (natural frequency 133 c./sec.) with a beating reed at the vertex. Keeping constant blowing pressure he gradually increased the length of the reed. The frequency of the system fell until the reed-length became 4·7 cm., at which the reed and first pipe overtone were in tune. Any slight further increase in length caused the note to drop suddenly to the fundamental of the cone, at which it remained until the reed-length exceeded 5 cm., when it again began to fall. On reversing the procedure the jump to the overtone occurred with a 5-cm. reed, producing a sort of hysteresis loop in the frequency : reed-length curve. (Incidentally, these jumps were first noticed by the brothers Weber over a century ago.)

One is taught to regard the clarinet as a cylindrical pipe with a reed at the mouth, having a timbre composed entirely of the odd harmonics, but partials of even order can in fact be detected in its spectrum. Ghosh develops the suggestion—not substantiated by direct observation, as far as the writer knows—that the reed never entirely closes the mouthpiece, so that the acoustic admittance of the mouth never becomes zero at any epoch. It is further assumed that the pressure inside the player's mouth remains constant while the

reed acts in the fashion of a valve to produce periodic variations of pressure in the embouchure of the instrument. When all side-holes are closed and an antinode lies at the open end, a formula for the frequency which has already been given elsewhere is derived :

$$\tan \omega l/c = -S\omega/c\kappa$$

(ω=pulsatance, c=velocity of sound, l=length, S=area of cross-section of pipe, and κ=conductivity of embouchure). After giving the first five solutions of this equation, which correspond roughly to the 1, 3, 5, 7, 9 members of the harmonic series, Ghosh obtains approximate equations for the motion of the reed, considered as a wedge-shaped bar clamped at one end. On account of the aerial vibrations, the pressure in the mouthpiece fluctuates cyclically about the value $P/(1+Skl)$, where P is the pressure inside the player's mouth. The motion of the reed will impose further pressure variations on these. If the chink opens at an epoch of maximum pressure, more air will flow in and accentuate the peak of pressure, but if the opening occurs at a minimum there will be a tendency to level out the fluctuation, i.e. to make an antinode of the mouthpiece. Theory shows that the best condition for maintaining the coupled system will be that the flow of air into the mouthpiece should precede the phase of maximum pressure therein by a quarter-period. Finally, the feeble even members in the spectrum are shown to be components in the reed's vibration forced upon the unwilling column of air.

Carrière [74] returns to an old problem which has intrigued him before—that of the singing flame. His experiments underline a point made some years ago by the writer, that stationary waves are set up in the gas-supply tube as well as in the air column containing the flame. He has illustrated this rather neatly by showing that if the flame is lit at the top of a supply tube of moderate length debouching from a gas reservoir, the flame will sing without the aid of the usual surrounding glass tube. The frequencies in this case are partials for the supply tube.

§ 4. ACOUSTIC IMPEDANCE: PROPERTIES OF MATERIALS

N. W. Robinson [75] continues his measurements of acoustic impedance. Using the same apparatus as in former experiments (sketched in *Reports*, 2, 179), he has now determined the conductivity of orifices (*a*) at the junction between two tubes in series or in parallel, (*b*) multiple thin orifices in parallel, (*c*) long orifices forming constrictions in a tube or terminating a tube. The method necessitates the measurement of the pressure amplitude in the neighbourhood of the orifice and, as the earlier experiments had shown that the records depended on the position of the mouth of the probe relative to the orifice, an investigation was first made of the shape of the sound-waves near a constriction in the tube. These results are exhibited in the form of equipotential lines, and show how the plane wave-fronts are curved in and out again on either side of the constriction. With a number of holes side by side in a disc fitting the tube, the equivalent conductivity was the sum of that of the separate holes, unless they were so close as to cause mutual interference of wave-fronts. When holes were put in parallel

with the main tube (side holes) the conductivity was almost the same as that of a hole of the same size bored in a plate stopping the end of the tube ; but holes forming constrictions in the tube have a conductivity twice those at the end or in the side, confirming an opinion of Bate's that holes with two sides exposed in a resonance column should be reckoned as having double effect. By using a series of discs from 3 mm. up to 3 cm. thick, with holes all of the same diameter bored through, Robinson was able to test the variation in conductivity with length of constriction. Provided this is short for the wave-length, λ, concerned, the simple formula $1/\kappa = 1/\kappa_0 + L/A$ applies, where κ_0 is the conductivity for the corresponding hole in the thin plate, A the area, and L the effective length of the boring. As long as L/λ does not exceed 0·07, L may be taken without much error as the actual length of the bore-hole, but as the wave-length gets shorter the effective L may be double the measured one. On the other hand, the measurements could not be relied on when the length of the constriction was small or when it approached $\lambda/4$. Bate and Wilson [76] have measured the end-corrections of truncated cones by the less accurate tuning-fork and resonance-tube apparatus. They find that whereas a cylinder has a constant end-correction, that of a cone is a function of pitch. In certain cones of large apical angle the end-correction was negative ; the terminal antinode at the wide end fell within the tube.

Sato [77] excites Helmholtz resonators with the sound from an organ pipe of constant intensity and explores the surrounding field with a Rayleigh disc in order to find out to what extent the resonator reacts upon the sound-field. For simplicity he supposes that plane waves impinge upon the opening of the resonator and that spherical waves are sent back. At resonance a velocity potential based on these considerations adequately describes the field. When the resonator is not tuned to the oncoming sound waves, the distortion of the free field is negligible.

Lindsay [78] gives a good general account of filters, including examples from his own researches. Besides dealing with the familiar infinite and finite filter theory, he shows the effect of terminating a filter with a finite impedance. Thus, suppose we attach a cylindrical tube to one end of a long high-pass filter. The attenuation over the low-frequency range will be unaffected ; but instead of this being followed by a level plateau above the cut-off frequency, the high level is broken up into a series of resonance peaks and anti-resonance crevasses due to the finial tube. These indentations are naturally reduced in size if the said tube is stuffed with absorbent material. In conjunction with Focke [79] he deals with analogous problems in the propagation of compressional waves through a solid filter structure consisting of a metal rod loaded with equidistant collars. Tests are made on a short length of this filter in which naturally the transmission band is marred by definite peaks, but the system behaves on the whole as a low-pass filter. In addition to the familiar mechanical and electrical analogues of these systems, Lindsay cites a new one, viz. the crystal metal lattice as a filter for the transmission of electrons. Waetzmann [80] uses a bridge for measuring the acoustic impedance of the ear-drum set as terminal impedance to one of the tubes forming an " arm " of the bridge. To get good reproducible results

the patient must hold his head very still. From the measured impedance on such a bridge it is simple to derive an " absorption coefficient of the ear-drum " whose value in most people rises from 20 per cent in the bass to 70 per cent at 1200 c./sec. and then falls a little ; but there are considerable variations between one person and another. (The measurements do not go above 1600 c./sec.)

Hartig and Swanson [81] discuss the effect of transverse waves in columns of air in relation to transmission and absorption measurements, being led thereto by recent work on analogous problems in the passage of electromagnetic waves through " dielectric radio-wave guides " and hollow metal tubes. The acoustic phenomenon manifested itself as an annoyance to the authors in an experimental tube intended for the measurement of impedance by the stationary (plane) wave method. Above a certain frequency, difficulty in obtaining uniformly spaced nodes and antinodes was experienced, and attributed to transverse acoustic waves. The solutions, involving Bessel functions, for these waves are derived, for which the equations are given in cylintrical co-ordinates. Sections of the stream pattern both across the tube and—for short sections—along it, corresponding to various phase conditions round and along the boundary, are given. The former are like the Chladni figures on a circular plate while the latter evoke memories of G. I. Taylor's vortices between concentric cylinders in

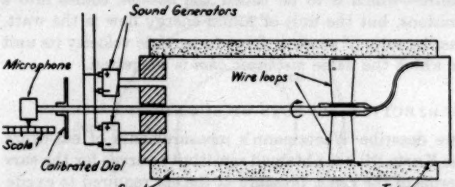


Figure 2 *a*.

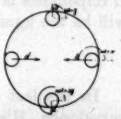


Figure 2 *b*.

relative motion. For the experiments a tube of galvanized iron 15 cm. diameter was enclosed in sand to ensure rigidity of its walls (figure 2 *a*). Loud-speakers were mounted at one end on the ends of two rectangular diameters, with various phase relationships between them (compare figure 2 *b*). A probe was caused to rotate about the axis at a distance which could be varied, and connected to a microphone so as to allow exploration of the pressure field, which is then compared with theory. Agreement is good considering that dissipation is neglected in the theory. Brillouin had already pointed out that, with suitable terminal devices for sending and receiving the transverse waves of a selected type, a high-pass filter could be constructed, since to every mode and diameter must correspond a critical frequency below which the given mode is theoretically forbidden. The authors expect that such filters would have a sharper cut-off than the usual type employing axial-wave transmission. In conclusion they utter a caution against the use of large ducts intended for plane-wave transmission

but which are in reality well adapted to the production of transverse waves, which will vitiate the results.

Taylor and Sherwin [82] have been using the method of the lined tube—" sound flue " they call it—for the absorption coefficients of acoustic materials, with the material applied to the sides as well as the ends to such purpose that sometimes the energy is completely absorbed by the time the sound has reached the far end, and stationary waves are precluded. Otherwise the coefficient is calculated from the relative amplitudes at maxima and minima in the pseudo-standing waves. Results for a material Acousti-celotex C1 and C3 are given, but cannot be compared with those by other methods since the attenuation is a function of the width of the flue. With high-pitched notes in the wider tubes it is doubtful, in view of the last paper, whether the waves are plane, and the variation with width of flue may be, we suggest, ascribed to radial waves. The method is one of the earliest in the history of architectural acoustics, having been used by Taylor in 1913.

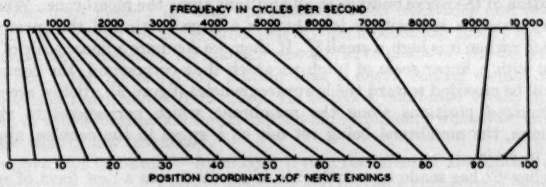
A paper by Kennelly and Cook [83] proposes a new system of units for electro-acoustics, the metre-kilogram-second system of Giorgi. This is homogeneous with practical electrical units already adopted by the International Electrotechnical Commission. One unit of the m.k.s. system we already commonly use, viz. the m./sec. for the velocity of sound. The m.k.s. unit of force—one joule per metre—which is to be called the *newton*, comes into a number of acoustic definitions, but the unit of sound energy flow is the watt. If impedance is defined as the ratio of acoustic force to particle velocity its unit will be the kgm./sec., for which the name *mechanic ohm* is suggested.

§ 5. SUBJECTIVE ASPECTS OF SOUND

In the last section we describe Waetzmann's measurements of ear-drum impedance. His colleague Kurtz [84] has obtained sensitivity curves for the ears of the same subjects in terms of the r.m.s. pressure at the ear required to excite the sensation of hearing. The two sets of data run hand in hand. As the " absorption coefficient " of the ear rises to a maximum near 1000 c./sec. the amplitude of pressure for the threshold of hearing diminishes in the same degree. Parker [85] has extended the work on sensitivity measurements with a standard source, which we noted in the last *Report*, to cover the effect on the threshold of fatiguing the ear with inexorable sound, and the recovery from this fatigue. For a few seconds after cessation of the fatiguing source the jaded ear shows results up to 10 db. greater than the normal threshold, but recovers in about a minute. The effect was a maximum when the test of audibility was made at the same pitch as the fatiguing stimulus, as one would expect, while if the two frequencies were sufficiently remote no loss of audibility was observed in the test. The results therefore support the idea of a localised stimulus in the sense of hearing. The range of frequency over which auditory fatigue is apparent increases with the strength of the stimulus, as also does the time taken to recover normal hearing.

Fletcher [86] has carried out masking experiments using the amplified

" thermal noise " due to statistical variation of the electrical potential difference across a conductor, resulting from the thermal agitation of its atoms. It is a convenient source because the amplification can be made to determine the intensity of the masking noise, while a filter will determine the limits of the band of frequencies which the noise covers. Within the cut-off limits of the filter the time average of the intensity can be taken as constant, so that the electrical power comprised in any frequency band is proportional to the width of that band. If now a thermal spectrum is impressed on the ear, suppose we wish to find the stimulation of the nerve patch at a distance x along the basilar



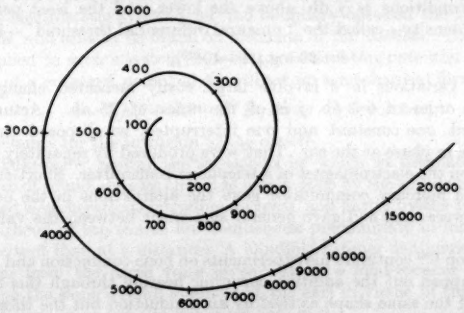


Figure 3.

membrane where there is resonance to the frequency f. The intensity δI in the small frequency band between f and $f+\delta f$ goes mainly to stimulate the patches between x and $x+\delta x$ (apart from a little spilling over to stimulate nerves outside this patch). When the threshold is reached the patch of nerves is excited at x where the maximum stimulation occurs. For convenience Fletcher defines a " patch " as " one per cent of all the nerve endings ", so that he has a hundred patches distributed along the basilar membrane. Let the actual power of agitation of a patch be J_x and let J_0 be the power of agitation for it

alone to produce the minimum audibility. It is assumed that the amount of masking can be taken as a measure of the agitation of a patch and that when the masking is the same for all frequencies all the patches have the same value of J_x/J_0. The intensity : frequency characteristic of an acoustic spectrogram which will do this has been experimentally found by the masking effect of a band δf over a pure tone of frequency f. The intensity falls between 100 and 4000 c./sec. and thereafter rises slightly to 10,000. The quantity J_x/J_0 is given by the requirement that, in terms of the arbitrary definition, x must be 100 when f is infinity. This spectrogram, then, is that of a noise which will produce constant stimulation of the nerve endings at all positions along the membrane. Wherever on this curve the intensity is low there is a large portion of the spectrum δf per patch ; when it is high, a small δf. If, then, we compare a linear scale of nerve patches with a linear scale of pitch to which they correspond, the connecting links will be crowded toward the lower frequencies (figure 3). Below are shown the estimated positions along the membrane which correspond to various frequencies, the membrane being set out as a spiral in the cochlea, approximately to scale.

Hughes [87] has made measurements of sensitivity to a new form of warble tone. The tone retains constant frequency but its intensity undergoes periodic changes to a lower value and back again. The minimum amplitude of fluctuation which can just be detected by the ear is determined. If the higher amplitude under these conditions is j db. above the lower, and the least perceptible difference of intensity—called the " uniaural differential threshold "—is δ, then

$$\delta = 20 \log_{10}(1+10^{j/20}),$$

so that small variations in δ involve large, easily measured changes in j. With δ of the order of 0·5 db. j is of the order of 25 db. Actually two notes were used, one constant and one interrupted, so disposed that their radiations were in phase at the ear. They were produced by separately exciting two windings on the electromagnet of a telephone transmitter. Short-circuiting one coil with a mercury commutator gave the alternations in the one note. Other devices were used and gave general agreement between the values of δ obtained.

N. A. Watson [88] continues his experiments on bone conduction and has now tentatively mapped out the audition curve for hearing through this channel. It is roughly of the same shape as that by air conduction but the threshold is higher, owing to losses in the head.

Reboul [89] likens the interior of the cochlea to a system of elementary condensers in parallel, with the liquid (*lymph*) forming the two conductors separated by the basilar membrane as dielectric. The potential difference between the fluid " plates " is supposed to be due to a difference of concentration. In his first paper he gives a theory of the propagation of sound-waves along the cochlea, considered as a conical pipe with elastic sides, and deduces the shape into which the membrane will be thrown by periodic disturbances coming through the round window, comparing his results with a model constructed by von Bekesy to illustrate this point of view, but which we believe was first

set up by G. Wilkinson. (This form of the resonance theory does not localise the response in the Helmholtz sense, since the shape of the membrane as a whole must determine the stimulation of the nerve endings.) Granted the validity of such a model, the passage of sound-waves and consequent deformation of the basilar membrane will set up potential variations in the " condenser ". The effect of a change of pressure on an elementary condenser is worked out and also that of a movement of the membrane. Whereas the former gives rise to potentials co-periodic with the oscillation, the latter set up one of double the mechanical frequency. As a result of comparisons of this theory with experience, the author lines himself up with opponents of both resonance and telephone theories ; the former on the ground that by his mechanical theory the basilar membrane as a response mechanism should have a lower cut-off at about 1000 c./sec. ; the latter on the ground that a nerve ending cannot respond to frequencies much above 1000 c./sec. He thinks a combined theory might be built up in which localisation would explain high-frequency sensation, and nerve response that of low frequency.

Arapova and Gersami [90] have been trying the effect of alternating electric currents on the sense of hearing generally and on the simultaneous listening to sound-waves in particular. In some cases frequency doubling from the A.C. occurred in the cochlea although the electric wave form was still simple-harmonic after passing through the body. In explanation they assume that there are electrostatic systems on the boundary between the lymph and the membrane, and cells to be found on elements of the cochlea. Condenser equations applied to such a system show that an alternating potential $E \sin \omega t$ in the presence of a constant potential E_0 will set up a mechanical force,

$$F = k(E_0^2 + 2E^2 + 2E E_0 \sin \omega t - \tfrac{1}{2}E^2 \cos 2\omega t),$$

which accounts for the intrusion of the octave.

Von Bekesy [91], testing the sensations by which we estimate the distance of a source of sound, decides that it is partly done by the smudging effect which echoes and reverberations introduce *en route*, and partly by the fact that in the neighbourhood of the source low frequencies predominate in most noises to a greater extent than at a distance. A blindfold listener distinguishes a whisper in his ear from the same voice given out by a loud-speaker at a distance (producing the same overall loudness at the ear) in virtue of the fact that the former is more clear-cut and has more bass in it.

Stout [92] has made oscillograph records of the vowel sounds *ah*, *oo*, and *ee* at various levels of pitch and loudness and analysed them to show how far the harmonic structure varies. Three frequencies were chosen, 147, 220 and 293 c./sec., covering the range of the adult male voice, but an attempt to select the same pair of intensity values for each pitch failed since it was impossible to sing vowels on the upper pitch at low intensity. Accordingly 4 and 14 db. were chosen for the two lower pitches, 14 and 22 db. for the uppermost. Three trained singers were selected as subjects. At the higher intensity greater energy was put into the spectrum above 1800 c./sec. and the enhancement of the treble was greater at the low pitch for *ah*, but at the high pitch for *ee*, and about the

same for all three in *oo*. There appeared to be no consistent change in the pitch location of the major intensity areas when the loudness increased. As pitch is raised at constant intensity, energy is shifted from the treble to the bass partials, and most of this goes into the fundamental.

Grutzmacher and Lottermoser [93] continue to make records of the melodic curves of speech, i.e. the variation in fundamental pitch when a series of words or syllables is spoken or sung. The same apparatus serves to measure the duration of " long " and " short " vowels and the mean pitch at which standard vowels are uttered. The latter show fair agreement between the four subjects whose efforts are recorded, but since the sounds are given neither in phonetic script nor in verbal examples, one cannot tell to which speech-sounds they relate. W. Trendelenburg [94] is also studying the singing voice, in particular the break in the registers. Not only is there a break between normal voice and falsetto when the vibrating length of the vocal cords changes, but the existence of other breaks is suspected and ascribed to coupling between the air in the larynx and that in the mouth, such as occur in a woodwind instrument as the length of the pipe is varied and the reaction upon the reed changes (cf. § 4). Since the vibrating length of the vocal cords is so small it is not suggested that coupling can affect them, but the air in the larynx is supposed to suffer the reaction. Tests of this theory are made by singing into a glass tube stuck in the mouth and over which the lips close so as to form a tripartite coupled system : larynx plus mouth plus artificial resonator. Either the vocal-cord sound is " held " as the tube is lengthened, or the tube is kept fixed while the unfortunate singer runs *portamento* up the scale. In each case records of the sound in the air are made. The glide tone shows distortions whenever it approaches one of the partial tones of the tube. (The variation is similar to that which refractive index suffers at anomalous dispersion.) In attempting to apply the results of these rather unnatural experiments to the unassisted voice, the author finds himself up against the difficulty that no such discontinuities occur when similar experiments are made without the tube except where the falsetto break comes, and this always at the same frequency. The other type of break when the tube is added should be a function of the size of the mouth resonator, and so occur at a different pitch for each vowel. He is therefore forced to the conclusion that the falsetto break is the only natural one. Any difficulties or uncertainties felt by the singer in placing his voice within the two registers must be purely muscular or psychological.

§ 6. TECHNOLOGY

In *architectural acoustics* interest is centred on the transmission of noise from place to place within a building, but there are one or two papers on reverberation. Thielscher [95] discusses references to the use of " acoustic vases " in the Greek Theatre and Early Christian Church, both in Vitruvius—the earliest writer on architectural acoustics—and in the descriptions of modern archaeologists. It seems that some ancient architects went further than the placing of vases in the proscenium and gave the walls a honeycomb structure, the unit cells of which were mistaken by later generations for ovens and even used for

this purpose when the edifice began to fall to ruins. In other cases a light curtain was placed in front of the stage, which held a handful of hidden choristers whose vocal output was raised by the vases to the level of that of fifty, thus anticipating the gramophone and loud-speakers which would be used to-day! The notion of putting resonators in the structure of the walls, with orifices communicating with the auditorium, recurs in the twentieth century to Zeller[96], who gives a theory for this practice, which is intended not to amplify the sound but to damp reverberation in virtue of the attenuation within the bodies of the resonators and in their necks. A few tests are described of walls honeycombed with resonators in this fashion.

Brillouin [97] describes a curious echo in a hall in Algiers where the space is divided by pillars into 25 squares (9 m.^{2}). Above each of these is a dome. When walking near the centre of any square, the visitor hears regular repeated echoes to each footstep. It appears that during the course of multiple reflections between each dome and the floor, the waves are focussed near the head of the listener. As each echo comes to a focus at a different instant, repeated echoes are heard.

Power [98] treats of a case of reverberation in which one surface only, the ceiling, of the room is highly absorbent and the rest hard. The use of Sabine's or Eyring's formula leads to results in disagreement with the measured reverberation time since they assume uniformly diffused sound energy and symmetrical disposition of the absorbents round the source. Power treats the special case by taking successive images of the source in the ceiling and working out the variation in energy density with time. Tests which gave good agreement with theory were made in offices with ceilings acoustically treated. In one case a room had a reverberation time of 0·96 sec. against a theoretical 0·98 sec., whereas Eyring's formula would have made it 0·57 sec.

P. E. Sabine has studied with Ramer [99] the effect of the average absorption of a chamber into which sound is being transmitted on measurements of the permeability of the intervening partition (of area A). Measurements are made of (a) the sound intensity on the incident side of the partition, I_0, (b) the intensity on the transmitted side of the partition, i.e. within the chamber, I_1. The transmission loss is $1/\tau$, where τ is given by $E_1 = \tau E_0 A/\alpha_1$, and involves α_1, the total equivalent absorption of the receiving room. Experiment bears out that τ does not depend on α_1, though a simple measurement of the difference of level I_0/I_1 on the two sides of the partition will not give the true transmission loss, since both A and α enter into this difference. The assumption that τ is independent of the *area* of the transmitting surface has still to be tested.

Constable [100] has made transmission experiments on partitions (in a block of one-room flats) and in water pipes. In the former case a loud-speaker was placed in one of the flats and the paths by which the sound reached the sixth flat on the same corridor were traced. Tests were also made in a vertical direction through the same ferro-concrete building. Because the floors were better insulators than the walls, a good deal of the noise came up by indirect paths. Along one floor, up and back along the upper floor was as important a path as one directly up through the ceiling of the first flat. In the other work

the author tried to prevent longitudinal vibrations being carried along water pipes by interposing rubber sections. One of the pipe lengths was driven at the end remote from the rubber inset over a range of warble frequencies and an electromagnetic detector of vibration in the further pipe section recorded the transmitted energy. The efficacy of the insulation increased very slowly with length after the first inch or two of rubber and its mean value diminished with rising frequency. Since the presence of water in the pipes made no appreciable difference and no resonance was observed at frequencies when it would be normally expected, it was suggested in the ensuing discussion that the author had not succeeded in setting up longitudinal vibrations in the water and tubes but only transverse vibrations in the metal portions of the latter.

In connection with *sound signalling*, Pierce and Noyes [101] have an interesting theoretical paper on the transmission of sound over a plane reflecting surface such as the sea. They carried out small-scale experiments using supersonic sources and about 300 wave-lengths separation of source and receiver. The signalling took place over a six-foot table covered in turn with aluminium, shallow water, textiles, sand, etc. Both sender and receiver could be raised above the table by equal amounts while the strength at the receiving station was noted with constant output. Owing to interference between direct and reflected waves, the intensity : height curves consist of distorted cycloids. With a good reflector such as water or metal the intensity drops to zero—complete interference—every few inches, but over soft "grounds" having a low internal velocity of sound, the undulations in the curve as the height is varied become less marked (like epicycloids). More complicated interference patterns resulted when a temperature gradient was set up in the air by warming the aluminium sheet. Acoustic mirages were formed and if the gradient were such as to bend the sound waves upward no signal got through until a critical height above the "ground" was attained. Very poor transmission at low angles over sand in sunlight results from a combination of these two effects, but an inverted temperature gradient—a belt of hot air over a cold sea—may increase the range of signalling. For signalling in liquids Klein [102] has catalogued the methods of measurement now available for amplitude in such media. "Laboratory measurements must be made in some form of standing wave system. Furthermore, experience has shown that for the determination of absolute sound in liquids it is more desirable to perfect and simplify the stationary wave pattern than to attempt to annihilate it. It is quite different from 'dead' room studies in air." That being granted it is rather surprising that the hot-wire anemometer is not mentioned, though the Rayleigh disc, acoustic radiometers and the diffraction of light beams (for supersonics) are described. In this connexion we can but draw attention to improvements in, and measurements of, the efficiency of the siren, pioneer of acoustic signalling sources, by Schiesser [103].

Sound reproducers have a number of works to their credit, three papers being on testing. Freygang [104] describes an apparatus for quickly obtaining the spectrogram of a loud-speaker at different points within a hall. These, compared with one made in a free sound-field, enable one to deduce the variation of absorption with pitch in the room. Thorsen [105] measures the amplification

afforded by deaf-aid appliances both objectively and as estimated by a person of normal hearing, and gets fair agreement between the two. Nahrgang [106] has constructed an elaborate harmonic tone generator working on the beat-frequency principle, and uses it for masking and timbre-change tests.

Two papers deal with gramophone pick-ups. Löfgren [107] describes the distortion which may be introduced by faulty orientation of the tone arm. Since the arm is usually pivoted at a point just outside the circumference of the disc, the pick-up has, strictly speaking, a correct setting—arm tangential to groove—for an unique value of the radial distance of the groove from the axis of the table. In practice this setting is obtained somewhere near the middle of the trace. At the beginning and end of the record the needle is dragging against one or other of the sides of the groove as the table turns. The analysis of the distortion factor in the simple case of a recorded sine wave is worked out for various radii of trace with respect to tone-arm radius, and then for several overlying S.H.M.s. The distortion has quite a different character to that of an overloaded valve amplifier, since higher summation tones intrude and give lack of fidelity to the high treble. It should be possible by careful attention to design to ensure greater purity than is at present tolerated. Pierce and Hunt [108] mention this form of distortion but consider that a more serious blemish is caused by the stylus not bedding down properly into the groove, so that the needle rides up the sides of the groove as this undulates from side to side or up and down during modulation. The ratio of groove and needle curvatures is the criterion for the upper limit of reproducible frequency. Equality in the radii corresponds to 40 per cent total distortion for the " hill and dale " type and 20 per cent for the " meandering " type of recording. The latter process generally gives greater fidelity but the purity varies more rapidly with the above distortion parameter. For best results they recommend lateral recording with a needle of small radius of curvature in a plane perpendicular to the groove.

Two papers deal with *airscrew noise*. The work of Ernsthausen [109] is with model propellers set up in a sound-proof " corner ", plotting the sound field round about with a spherical condenser microphone. Most of the vortex noise goes off in the plane of the propeller disc, but along the axis it is overpowered by the noise of the flexural vibrations of the blades. Hilton [110] has obtained some very clear photographs of the pressure waves set up by the blades moving at speeds in the neighbourhood of the velocity of sound. There is no discontinuity in noise intensity as the tip-speed passes through this value, since the intensity remains directly proportional to the speed of revolution, but the quality changes from one with a prominent fundamental to a note harsh and richer in overtones at a critical speed when shock waves are first registered on the photographs. The existence of shock waves suggests a rapid rise of pressure across the wave-front, the Fourier analysis of which would yield many overtones. It seems that at low speeds the excess pressure in the sound wave is spread out over a band whose width is about half a wave-length, but that at a critical speed the wave-front condenses into a single line retaining the same total amplitude. This critical speed is probably that at which the lift and drag coefficients of an aerofoil show abrupt transitions from their low-speed values,

due to the local velocity of sound being exceeded at some point along the aerofoil. The noise is concentrated into compressions and rarefactions having the same frequency as the number of air-screw blades passing a given point per second.

§ 7. ELECTRICAL INSTRUMENTS OF MUSIC

An international committee, sitting under the presidency of Dr G. W. C. Kaye [111] and fostered by the British Institute of Standards, has recently decided on 440 c./sec. to be the standard A for all broadcasting, and is now considering how the standard can be adhered to. The wind instruments are of course the chief offenders in their variability with atmospheric temperature, and it is rather surprising that the oboe has been allowed to retain its position as arbiter of orchestral pitch for so long. It will not be easy for military bands in tropical climates, for example, to adhere to the standard pitch on all occasions, but attention is being directed to electrical tone producers both for giving out the standard frequency and even for producing the whole of the music, since they are remarkably free from temperature variation and readily lend themselves to the synthesis of any desired quality.

Some of these are modifications of existing instruments and are to be classed as electro-acoustical, while others are more completely electric and have no external similarity with conventional musical instruments. In America the generic name of "electronic" is given to them, though this is to be understood as implying no more than that valve amplifiers or oscillators appear somewhere in the equipment.

Electro-acoustic instruments. The least ambitious of these replace the sound-board of a conventional stringed instrument by pick-ups (one or more to each string) whose unmodulated electrical vibrations are amplified and delivered to a loud-speaker. Such are the Neo-Bechstein, a pianoforte without soundboard, and the Electric Guitar, which have pick-ups behind the strings. On all these instruments a magnetic pick-up may be used, that is to say, as the steel wire vibrates it changes the magnetic induction in a small coil nearby, or it may be an electrostatic pick-up in which case the vibrator moves to and from a plate behind the wire. The wire is then charged to a difference of potential over the plate so that the pair make up a condenser of varying capacity, which can be connected into the grid circuit of a valve amplifier.

The station of the pick-up along the wire influences the quality of the vibration passed on to the amplifier. Not only this, but the quality of the note of the wire itself exerts a powerful influence, and this in turn—according to well-known acoustical laws—is determined by the position and extent of the agent which sets the wire in vibration. Thus if the string is struck as in the pianoforte there will be a greater number of harmonics in the resulting tone if the hammer is sharply pointed and hard than if it is rounded and resilient. Furthermore, the choice of striking point operates in this way; that to each harmonic in the complex sound of the vibrating string there is a series of nodes and antinodes. If the location of the striking point is such that one or more of the harmonics require a node at that point then these particular harmonics

are absent in the resulting note. Similar principles apply to the situation of the pick-up. Electromagnetic oscillations which it delivers to the amplifier will be deficient in those harmonics which have a node on the string opposite the pick-up, while those with antinodes there will receive preferential treatment. For this reason some inventors use a number of pick-ups to each wire and mix the resulting tones, though this is an expensive method, and it is just as easy to modulate the tone from a single pick-up in any desired way.

Other instruments of the same type use reeds as the acoustic vibrators (Miessner [115]). Sometimes these are excited by hammers in the manner of a celeste (tuning-fork pianoforte); at other times wind-maintained reeds are employed as on the harmonium.

Instruments for aether music. Under this ambitious heading—following the inventors—we may group those instruments which derive music directly from tuned circuits in oscillation and have no mechanical vibrators other than the diaphragm in the ultimate loud-speaker, if any. The prototype of these was the singing arc of Duddell. The arc was shunted by a coil and condenser whose inductance and capacity determined the natural frequency of the circuit. By continuously variable or step by step changes in inductance it was possible to play tunes on the arc. Since then numerous instruments using valve oscillators (audions) have been patented, but those with keyboards and one oscillator to each note have proved too expensive both in first cost and upkeep of their numerous valves to gain a footing in the market. There remain the solo instruments, of which the Theremin and the Trautonium have reached commercial production. Theremin's instrument, which was demonstrated in this country about ten years ago before a distinguished audience from the artistic world under the aegis of Mr C. B. Cochran, is a beat-frequency oscillator in which the capacity controlling the frequency consists of a copper loop and a baton held in the hand (or sometimes the hand itself). The beat-note when the hand is away from the loop is above the audible limit but descends through the musical gamut as the hand approaches the loop in virtue of the increasing capacity, corresponding to the approach of an earthed conductor to an insulated plate in a familiar experiment of the elementary physics course. With this mode of control the music is confined to a series of single notes with a *glissando* between each which—together with the constant "feeling" for the note, like a horn-player burbling for the harmonics of his instrument—becomes rather distressing to the musical ear, in spite of the undoubted skill in performance of the inventor.

In the Trautonium, sponsored by the Telefunken Company of Germany, *portamento* playing is still possible, or one may proceed directly from note to note in proper musical style. A grid-glow tube is used as a variable-frequency generator. The grid potential which determines the pitch is controlled by the length of resistance wire cut off by the player pressing some point on the wire on to a mark on a metal plate behind the wire, like the frets on a banjo. Another resistance under the plate is varied by the *pressure* of the finger on the plate and so alters the loudness of the sound. It is therefore possible to get as nice a variation in playing as a virtuoso may get from a violin by the

mere graduation in the position and pressure of a single finger on the wire, including effects such as the *vibrato* and *staccato* playing.

Electric organs. The Telharmonium of Cahill antedated the valve amplifier and speaker. His apparatus, like those of the present day, produced electric currents of varying frequency and intensity from rotating elements. These currents were passed over the United States Telephone System and provided music for the household subscriber in a service very similar to that now available to subscribers in this country by the service operated by the British Broadcasting Corporation in conjunction with the Post Office Telephone System. However, after the equipment had been built, the experiment had to be abandoned owing to interference and inductive effects on other lines in the far from perfect telephone system of those days (1906), though his system of production contained all the essentials of an electrical organ. There are now a number of these instruments to choose from and at least three in Great Britain, of which the Hammond (an American invention) is probably the most widely diffused.

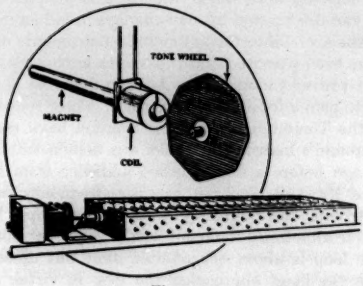


Figure 4.

A synchronous motor drives a series of 91 tone generators through gears and pinions. One of these is shown in figure 4. The tone wheel is a polygonal plate about the size of half-a-crown, rotating near a permanent magnet on which a coil is wound. As a high point on the wheel passes the magnet it induces a pulse of current in the coil. The speeds of rotation and the number of corners on each wheel are so calculated that each disc produces one of the 91 partials used in building up the fundamentals and required timbre. The latter is decided in the harmonic controller wherein the various frequencies are superimposed and flow as a single complex electrical wave to the pre-amplifier located in the console. There is of course an intensity control for each stop and a "swell pedal" which alters the overall intensity between the pre-amplifier and the loud-speakers.

The other type uses electrostatic production of tone. This is the principle employed in the Compton and the Midgley electrophonic instruments. Two

electrodes are spaced a small distance apart in the air, one moving relatively to the other. The two members constitute two plates of a condenser, usually at a fixed distance apart. An undulating variation in capacity between the members is caused by rotating one member on which sinusoidal grooves are inscribed, while the other is kept fixed. As this system is liable to default due to buckling of one of the plates or lack of parallelism, Midgley keeps both condenser plates fixed, while varying the dielectric constant. One of the stationary discs of insulating material, to the face of which are fixed eight concentric rings of conducting material cut into sine waves, is shown on the left of figure 5. The number of forms to each ring runs in powers of 2 from the centre to the outer ring, which has 256 undulations. A pair of these stationary discs face each other, while between them rotates another disc of bakelite, seen on the right of figure 5. The rotating disc is provided with eight concentric rings of apertures, each of which covers a half wave-length of the corresponding

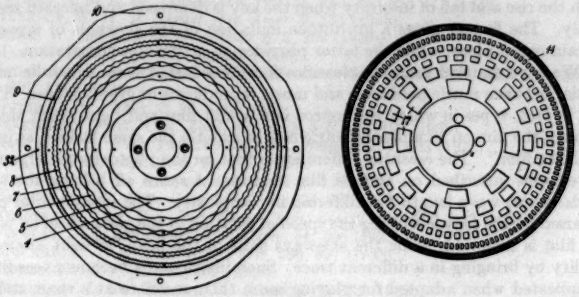


Figure 5.

sine wave. These condensers, of varying capacity, are connected in the grid circuits of valves so as to produce the necessary fundamentals and overtones for covering the musical gamut.

On most of these instruments the control of timbre is assured by stops as on the classical organ, in some of which the timbre is pre-selected to give close imitations of common organ stops or orchestral instruments. Other combinations lie at the choice of the player. Thus on the Hammond organ there is at each end a small manual of six or seven coloured keys, which determine the relative magnitude of the first six or seven harmonics of the fundamental of each note in the selective stops. Each of these control keys can be depressed through eight stages of intensity. Thus in the combination 6, 0, 7, 0, 3, 3, in which the numbers denote the relative intensities of the partials on the harmonic scale, the predominance of the odd harmonics point to a clarinet quality. A similar combination, with less of the fifth and sixth partials, would imitate a stopped flue pipe.

Although the three organs described normally give tones of fixed intensity, it is not difficult, though of course it adds to the elaboration, to introduce circuits which will damp the sound of each note from its inception and so mimic pianoforte, harp or bell tone, but of course the artificial wave form does not vary during attenuation in the same way that it does in some of these instruments. It is not beyond the wit of man to devise circuits which will change the quality at the same time as they make the note die away. Some inventors indeed derive both constant and evanescent tones, as required, from the same damped source, e.g. a struck reed. As soon as the sound has reached maximum intensity the pick-up is switched over to a circuit with a time constant which can be made to give long or short damping at will, or to maintain the tone without dissipation as long as the key is held down, in spite of the damping in the acoustic vibrator (this necessitates a negative resistance in the circuit). On the Midgley organ low-pass filters for preventing the high-frequency oscillations produced by the key contacts from reaching the amplifiers produce a lag in both the rise and fall of intensity when the key is depressed and released respectively. The former gives a gratuitous imitation of the slowness of speech of certain organ pipes, while the latter provides an artificial reverberation.

We have omitted from our classification those instruments which lie on the border between music producers and music reproducers or gramophones. There are, in fact, types in which the source is a sound film and light beam shining through the film on to a photoelectric cell, in exactly the same fashion as in the "talking film". The essential difference is that on the photoelectric organ the wave-forms inscribed on endless film comprise a series all having the same fundamental wave-length but differing in complexity. The fundamental pitch is varied by the player altering the speed of rotation of the glass wheel on which the film is mounted, while the sideways shift of the beam of light alters the quality by bringing in a different trace. Such instruments become exceedingly complicated when adapted for playing more than one note at a time, and are scarcely likely to find favour save where talking-film projector equipment is already installed. Even then it may be cheaper to purchase a library of sound films for the music required than to employ a photoelectric organ and organist.

We have spoken so far of the electrical keyboard instrument as replacing the pianoforte and organ. This it can do quite well, since it is the aim of the makers of such instruments to secure uniformity of quality throughout the pianoforte and through each stop of pipes or reeds on the organ. When one considers how far it is feasible to replace an orchestral instrument with an electric tone producer, another problem arises. One must consider what are the characteristic "formants" of the instrument. A solo instrument of extensive range does not have the same wave-form throughout. That of the clarinet, for example, differs markedly in the upper register from the lower, while the timbre of the violin A played on the G string is not the same as that of the open A string. Each individual instrument has, in fact, characteristic resonances within its structure which reinforce certain partial tones in any note which is played and suppress others. No pre-set "mixtures" such as the Hammond organ has could give a satisfactory imitation of the clarinet or violoncello. To do this

would necessitate the addition of elaborate filters having a fine structure of resonances to imitate the complex formants originating in the case and air cavities of these instruments.

REFERENCES

(1) Kukkamaki, T. J. *Ann. Phys., Lpz.*, **31**, 398 (1938).
(2) Waetzmann, E., Scholz, J. and Kruger, H. *Akust. Z.* **3**, 241 (1938).
(3) Colwell, R. C., Friend, H. W. and McGraw, D. A. *J. Franklin Inst.* **225**, 579 (1938).
(4) Herzfeld, K. F. *Phys. Rev.* **53**, 899 (1938).
(5) Santon, H. L. *J. Chem. Phys.* **6**, 30 (1938).
(6) van Itterbeek, A., van Paemel, O. and Thys, L. *Physica*, **5**, 153, 298, 593, 640, 845 and 889 (1938).
(7) Schmidtmuller, N. *Akust. Z.* **3**, 114 (1938).
(8) Knudsen, V. O. and Fricke, E. F. *J. Acoust. Soc. Amer.* **10**, 89 (1938).
(9) Kuchler, L. *Z. phys. Chem.* **41 B**, 199 (1938).
(10) Metter, I. M. *Acta Physicochem.* **9**, 845 (1938).
(11) Jatkar, S. K. K. *J. Indian Inst. Sci.* **21**, 245, 455 and 477 (1938).
(12) Hulswit, W. H. *J. Acoust. Soc. Amer.* **9**, 224 (1938).
(13) Burk, W. and Lichte, H. *Akust. Z.* **3**, 259 (1938).
(14) May, J. *Proc. Phys. Soc.* **50**, 553 (1938).
(15) Salceanu, C. and Istraty, C. *C.R. Acad. Sci., Paris*, **206**, 329 and 502; **207**, 1184 (1938).
(16) Dutta, A. K. *Phys. Z.* **39**, 186 (1938).
(17) Spakowsky, P. G. *C.R. Acad. Sci., U.S.S.R.*, **18**, 169 (1938).
(18) Rao, B. V. R. *Proc. Indian Acad. Sci., Bangalore*, **7**, 163 (1938).
(19) Bär, R. *Helv. phys. Acta*, **11**, 472 (1938).
(20) Liepmann, H. W. *Helv. phys. Acta*, **11**, 381 (1938).
(21) Hirschlaff, E. *Proc. Camb. Phil. Soc.* **34**, 296 (1938).
(22) Yosioka, K. *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **806**, 843 (1938).
(23) Findlay, J. C. and others. *Phys. Rev.* **54**, 506 (1938).
(24) Biquard, P. *C.R. Acad. Sci., Paris*, **206**, 897 (1938).
(25) Seifen, N. *Z. Phys.* **108**, 681 (1938).
(26) Bergmann, L. and Oertel, H. *Akust. Z.* **3**, 332 (1938).
(27) Grobe, H. *Phys. Z.* **39**, 333 (1938).
(28) Bazulin, P. *C.R. Acad. Sci., U.S.S.R.*, **19**, 153 (1938).
(29) Buss, W. *Ann. Phys. Lpz.*, **33**, 143 (1938).
(30) Cance, J. P. *J. Phys. Radium*, **9**, 308 (1938).
(31) Kretschmer, S. I. and Rschevkin, S. N. *C.R. Acad. Sci., U.S.S.R.*, **20**, 17 (1938).
(32) Lucas, R. *C.R. Acad. Sci., Paris*, **206**, 827 (1938).
(33) Hoesch, K. H. *Z. Phys.* **109**, 606 (1938).
(34) Laidler, T. J. and Richardson, E. G. *J. Acoust. Soc. Amer.* **9**, 217 (1938).
(35) Freundlich, H. and Gillings, D. W. *Trans. Faraday Soc.* **34**, 649 (1938).
(36) Thieme, E. *Phys. Z.* **39**, 384 (1938).
(37) Söllner, K. *Trans. Faraday Soc.* **34**, 1170 (1938).
(38) Hermans, J. J. *Phil. Mag.* **25**, 426, and **26**, 674 (1938).
(39) Meyer, E. and Buchmann, G. *Akust. Z.* **3**, 132 (1938).
(40) Fox, F. E. and Rock, G. D. *Rev. Sci. Instrum.* **9**, 341 (1938); *Phys. Rev.* **54**, 223 (1938).
(41) Muller, G. G., Black, R. and Davis, T. E. *J. Acoust. Soc. Amer.* **10**, 6 (1938).
(42) Bate, A. E. *Proc. Phys. Soc.* **50**, 293 (1938).
(43) Waetzmann, E. and Kurtz, R. *Ann. Phys., Lpz.*, **31**, 661 (1938).
(44) Schuck, O. H. *J. Acoust. Soc. Amer.* **10**, 119 (1938).
(45) Essen, L. *Proc. Phys. Soc.* **50**, 413 (1938).

(46) Ni Tsi Ze and Chung Ming-Sang. *J. Phys. Radium*, **9**, 52 (1938).
(47) Zacek, A. and Ptrzilka, V. *Phil. Mag.* **25**, 164 (1938).
(48) Pierce, G. W. and Noyes, A. *J. Acoust. Soc. Amer.* **9**, 185, 205 (1938).
(49) Focke, A., Lindsay, R. B. and Wilks, C. R. *J. Acoust. Soc. Amer.* **9**, 348 (1938).
(50) Baumgardt, E. E. *C.R. Acad. Sci., Paris*, **206**, 1284 ; **207**, 273 (1938).
(51) Walti, A. *Helv. phys. Acta*, **11**, 113 (1938).
(52) Reissner, H. *Helv. phys. Acta*, **11**, 140 (1938).
(53) Bär, R. *Helv. phys. Acta*, **11**, 397 (1938).
(54) Levi, F. and Nath, N. S. N. *Helv. phys. Acta*, **11**, 408 (1938).
(55) Pierce, G. W. and Noyes, A. *J Acoust. Soc. Amer.* **9**, 301 (1938).
(56) Allen, G. E. *Phil. Mag.* **26**, 609 (1938).
(57) Landon, D. H. and Davies, R. M. *Phil. Mag.* **26**, 816 (1938).
(58) Jarnak, P. *J. Franklin. Inst.* **225**, 315 (1938).
(59) Meinel, H. *Z. tech. Phys.* **19**, 297 and 421 (1938).
(60) Ghosh, M. *Indian J. Phys.* **12**, 317 and 437 (1938).
(61) Waller, Mary D. *Proc. Phys. Soc.* **50**, 70 (1938).
(62) Colwell, R. C., Friend, A. W. and Stewart, J. K. *J. Acoust. Soc. Amer.* **10**, 68 (1938).
(63) Arts, J. *J. Acoust. Soc. Amer.* **9**, 344 (1938).
(64) Rao, K. N. *Proc. Indian Acad. Sci., Bangalore*, **7**, 75 (1938).
(65) Holle, W. *Akust. Z.* **3**, 321 (1938).
(66) Coop, J. J. *J. Acoust. Soc. Amer.* **9**, 321 (1938).
(67) Suzuki, S. and Miduno, Z. *Proc. Phys.-Math. Soc. Japan*, **20**, 926 (1938).
(68) Brown, G. B. *Nature, Lond.*, **141**, 11 (1938) ; *Sci. Progr.* **33**, 129 (1938).
(69) Trendelenburg, F., Thienhaus, E. and Franz, E. *Akust. Z.* **3**, 7 (1938).
(70) Bonar, C. P. *J. Acoust. Soc. Amer.* **10**, 32 (1938).
(71) Mokhtar, M. *Trans. Durham Phil. Soc.* **9**, 352 (1938).
(72) Auger, L. *C.R. Acad. Sci., Paris*, **207**, 216 (1938).
(73) Ghosh, R. N. *J. Acoust. Soc. Amer.* **9**, 255 (1938).
(74) Carrière, Z. *Rev. d'Acoust.* **7**, 20 (1938).
(75) Robinson, N. W. *Proc. Phys. Soc.* **50**, 599 (1938).
(76) Bate, A. E. and Wilson, E. T. *Phil. Mag.* **26**, 752 (1938).
(77) Sato, K. *Proc. Phys.-Math. Soc. Japan*, **20**, 312 (1938).
(78) Lindsay, R. B. *J. Appl. Phys.* **9**, 612 (1938).
(79) Lindsay, R. B. and Focke, A. B. *J. Acoust. Soc. Amer.* **10**, 41 (1938).
(80) Waetzmann, E. *Akust. Z.* **3**, 1 (1938).
(81) Hartig, H. E. and Swanson, C. E. *Phys. Rev.* **54**, 618 (1938).
(82) Taylor, H. O. and Sherwin, C. W. *J. Acoust. Soc. Amer.* **9**, 331 (1938).
(83) Kennelly, A. E. and Cook, J. H. *J. Acoust. Soc. Amer.* **9**, 336 (1938).
(84) Kurtz, R. *Akust. Z.* **3**, 74 (1938).
(85) Parker, R. C. *Proc. Phys. Soc.* **50**, 108 (1938).
(86) Fletcher, H. *J. Acoust. Soc. Amer.* **9**, 275 (1938) ; *Proc. Nat. Acad. Sci., Wash.*, **24**, 265 (1938).
(87) Hughes, J. W. *Phil. Mag.* **26**, 635 (1938).
(88) Watson, N. A. *J. Acoust. Soc. Amer.* **9**, 294 (1938).
(89) Reboul, J. A. *J. Phys. Radium*, **9**, 185 and 428 (1938).
(90) Arapova, A. and Gersami, G. *Z. tech. Phys., U.S.S.R.*, **5**, 447 (1938).
(91) Bekesy, G. von. *Akust. Z.* **3**, 21 (1938).
(92) Stout, B. *J. Acoust. Soc. Amer.* **10**, 137 (1938).
(93) Grutzmacher, M. and Lottermoser, W. *Akust. Z.* **3**, 183 (1938).
(94) Trendelenburg, W. *S.B. preuss. Akad. Wiss.* 1 and 21 (1938).
(95) Thielscher, P. *Akust. Z.* **3**, 93 (1938).
(96) Zeller, W. *Akust. Z.* **3**, 32 (1938).
(97) Brillouin, J. *Rev. d'Acoust.* **7**, 117 (1938).

(98) Power, J. P. *J. Acoust. Soc. Amer.* **10**, 98 (1938).
(99) Sabine, P. E. and Ramer, L. G. *J. Acoust. Soc. Amer.* **10**, 102 (1938).
(100) Constable, J. E. R. *Proc. Phys. Soc.* **50**, 360 and 368 (1938).
(101) Pierce, G. W. and Noyes, A. *J. Acoust. Soc. Amer.* **9**, 193 (1938).
(102) Klein, E. *J. Acoust. Soc. Amer.* **10**, 105 (1938).
(103) Schiesser, H. *Akust. Z.* **3**, 363 (1938).
(104) Freygang, H. G. *Akust. Z.* **3**, 80 (1938).
(105) Thorsen, V. *Akust. Z.* **3**, 218 (1938).
(106) Nahrgang, S. *Akust. Z.* **3**, 284 (1938).
(107) Löfgren, E. *Akust. Z.* **3**, 350 (1938).
(108) Pierce, J. A. and Hunt, L. V. *J. Acoust. Soc. Amer.* **10**, 14 (1938).
(109) Ernsthausen, W. *Akust. Z.* **3**, 141 (1938).
(110) Hilton, W. F. *Proc. Roy. Soc.* **A, 169**, 174 (1938).
(111) Kaye, G. W. C. *Nature, Lond.*, **142**, 820 (1938)
(112) Vierling, O. *Z. Ver. dtsch. Ing.* **76**, 625 (1932).
(113) Janofsky, W. *Elektrotech., Z.* **54**, 675 (1933).
(114) Aschoff, *Naturwissenschaften*, **25**, 598 (1937).
(115) Meissner, B. F. *Proc. Inst. Radio Engrs, N.Y.*, **24**, 1427 (1936).
(116) Galpin, F. W. *Proc. Musical Assoc.* **64**, 71 (1938).

FLUID MOTION

TRANSITION FROM LAMINAR TO TURBULENT FLOW IN A BOUNDARY LAYER

By A. FAGE, A.R.C.S., F.R.Ae.S.

When a fluid of small viscosity like air or water flows past a fixed solid body, the velocity rises rapidly from zero at the surface to the value in the main stream (except when this value is very low), and this rise takes place within a thin layer of fluid adjacent to the surface. This thin layer of fluid is called the *boundary layer*. The effect of viscosity is important only in this layer, and the effect in the main stream outside the layer can be ignored. Within the layer the velocity gradient is large, and however small the viscosity may be, the tangential stresses within the layer exert a considerable influence upon its motion. The boundary layer at the front of a body is extremely thin, and as the fluid flows along the surface the retarding effect of the tangential stresses continually increases the width of the layer. The motion of the fluid within the layer is controlled partly by dynamic forces associated with the kinetic energy and pressure of the fluid, and partly by the tangential stresses associated with the viscosity. The tangential stresses are most effective near the surface, and they become less important as the distance from the surface increases, finally becoming negligible at and beyond the outer limit of the layer. The tangential stresses do work at the expense of the total pressure energy, and the boundary layer can be taken to be the thin layer of fluid next to the surface within which a loss of total pressure * can be measured, the total pressure in the stream outside the layer being taken as datum.

The motion in a boundary layer at the front of a body is laminar, but this laminar motion breaks down and becomes turbulent if the distance which the layer has traversed along the surface, or the velocity in the main stream outside the layer, is large enough. This breakdown from laminar to turbulent motion is not an abrupt but a gradual process extending over a definite region on the surface, and often this region is not stationary but traverses to and fro on the surface, the extent of the traverse depending mainly on and increasing with the disturbances in the main stream outside the layer. Transfer of momentum across a laminar layer is by molecular action and across a turbulent layer by molar action, and the characteristic velocity distributions across a layer for laminar and turbulent motions are different. It should be mentioned that for practical purposes it is customary to consider the motion in a boundary layer to be laminar, even when fluctuations of velocity exist in the layer, provided that these fluctuations are of such a character that the measured distribution of mean velocity across the layer closely resembles the distribution calculated on the assumption of steady flow. A change in

* The total pressure is the pressure measured by a pitot tube facing the stream.

the type of the velocity distribution across a boundary layer occurs on passing through a transition region. In addition there is also a rapid increase in the thickness of the layer. An idea of the magnitude of this increase in thickness is shown in figure 1, which gives the thickness of the boundary layer on a long cylindrical body whose cross-section has the shape of a symmetrical aerofoil. The thickness of the boundary layer is drawn true to scale from measurements taken on a model aerofoil of 40 in. chord in an air stream having a velocity of 80 ft. per sec. It will be seen that the layer is very thin at the nose—the measured thickness is 0·001 chord at 0·05 chord from the nose—and increases in thickness towards the tail. Transition occurs at about 0·2 chord from the nose, and at this position (marked T in the figure) there is a rapid increase in the thickness of the layer. The thickness of the laminar layer immediately before transition is about 0·0022 chord.

Transition phenomena have recently been studied with marked success, particularly in connexion with problems of aeroplane drag. The drag of an aeroplane wing depends appreciably on the position of transition from laminar to turbulent flow in the boundary layer on its surface, because of the increase in the frictional intensity on the surface, and an idea of the nature of this dependence can be gathered from the fact that a saving of about

Figure 1.

30 per cent in drag can be achieved by a shift of the position of transition from the leading edge of a wing to a distance of 0·3 of the chord behind the leading edge. The wing drag of a modern aeroplane is about 50 per cent of the total drag, so that reduction of wing drag is a matter of prime importance, and the most promising method of reduction is by the use of wing sections for which the transition occurs well beyond the leading edge. It is for this reason that many researches on the nature of transition from laminar to turbulent flow in a boundary layer and on the factors which influence transition have been made in aeronautical laboratories. The present report deals almost exclusively with these researches. Apart from the application to aeronautical problems, the results of these researches have a wider application in the many domains of fluid dynamics involving relative movement of solids and fluids.

§ 1. MEASUREMENT OF THE POSITION OF TRANSITION POINT

The transition from laminar to turbulent flow in a boundary layer * does not take place at a point but over a length of the surface, and this length on a body in a stream of low turbulence is small. The point where the transition begins, known as the transition point, can be measured with good accuracy.

* We are here considering the flow in a section normal to the surface and in the general direction of flow.

An essential feature of boundary-layer flow immediately beyond the transition point is a rapid thickening of the layer, and the accompanying change in the velocity profile is such that the velocity very near the surface increases appreciably. The transition point can therefore be easily located from measurements of total pressure taken with a small flattened pitot tube in contact with the surface [1]–[4]. As the tube (P_1 in figure 2) is moved along the surface from the laminar layer into the transition region there is a slow fall of pressure in the laminar layer and this is followed by a marked rise of pressure on passing into the transition region. The point where the rise begins is taken to be the transition point. Beyond the transition region, where turbulent flow is fully developed, the pressure falls with a movement of the tube along the surface. An alternative method of location [1]–[3], commonly used on aeroplane wings in

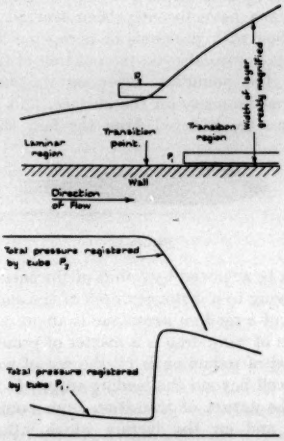


Figure 2.

flight, depends on the detection of the thickening of the layer from the reading of a small pitot tube (P_2 in the figure) traversed parallel to the surface at a distance just greater than the thickness of the laminar layer. The full total pressure in the stream outside the layer is registered before the tube enters the transition region, whereas a fall is registered when the tube enters the thickening transition layer (see curve of figure 2). Tubes having larger bores can be used in this alternative method, so that in flight experiments the lag effect on the pressure registered, arising from pressure and temperature changes in the atmosphere with altitude, is reduced. Although the thickness of a laminar layer just before transition on a smooth aeroplane wing at ordinary speeds of flight does not exceed about 0·05 in., the preliminary estimate of the distance

from the surface at which the tube should be placed need not be very accurate, because the thickness of the layer increases rapidly beyond the transition point.

Complete surveys of the flow in a transition region have been made in aeronautical laboratories, with small pitot tubes [4], [5] and with small hot-wire anemometers [6]–[10], and maps giving contour lines of equal speed constructed from such surveys [7] illustrate essential features associated with the flow in and near a transition region. Indications of the onset of transition can also be revealed by oscillograms obtained with a hot-wire anemometer recording on the screen of a cathode-ray tube [10], [11] or an Einthoven galvanometer [12], by the behaviour of a filament band of smoke in a boundary layer [13], and by the onset of unsteadiness in the pressure given by a pitot tube registering on a lightly damped gauge. Further, it has often been observed [25] that there is a local fall of static pressure, due to the rapid increase of boundary-layer thickness, in the immediate neighbourhood of transition.

The position at which transition in a boundary layer of a body occurs, depends on the velocity of the main stream (or more strictly on Reynolds number), the intensity and scale of turbulence in the main stream, the pressure gradient in the direction of flow, and finally on the curvature and roughness of the surface. Information on the dependence of transition on these factors has been obtained from investigations on aeroplane wings in flight and on bodies mounted in wind-tunnel streams, and it is now proposed to consider in some detail the influence of each of these factors.

§2. EFFECT OF TURBULENCE ON TRANSITION

The effect of turbulence on transition is most conveniently studied on a body immersed in the stream of a wind tunnel. Wind tunnels having streams substantially free from turbulence can be obtained by allowing air to flow through a fine-mesh honeycomb, a long settling chamber, and a large contraction before reaching the working chamber: and into such streams turbulence of known intensity and scale can be introduced by a grid mounted normal to the direction of flow. At a sufficient distance behind a grid, and indeed in most wind tunnels without a grid, the turbulence tends to be isotropic. The specification and measurement of this turbulence have to be briefly considered before the effect of turbulence on transition is discussed. Turbulence is the irregular random fluctuation with time of the speed at any point in the stream, and the intensity of turbulence is specified by the statistical root-mean-square values of the components of the deviations of the instantaneous speed from the mean value. The root-mean-square value of the component in the mean direction of flow, commonly denoted by $\sqrt{\overline{u^2}}$, can be measured with a hot-wire anemometer [14]–[16]. Platinum wires of diameter 0·0001 in. and length 0·05 in., obtained by removing the silver coating from Wollaston wire by a suitable process of etching and electrolysis in a bubble of nitric acid, are used. These fine wires behave satisfactorily up to air speeds of 80 ft. per sec.; and more robust wires of diameter 0·00025 in. have given good service in flight

up to a speed of 200 miles per hour. The variations of the potential drop across a wire, caused by the speed fluctuations, are amplified and applied to a thermal-type milliammeter, and from the recorded intensity of the fluctuating current output, the root-mean-square of the speed fluctuation is deduced. Compensation for the lag of a wire in response to rapid speed fluctuations is obtained by a suitable electrical network, consisting of an inductance and an adjustable resistance in series with a large fixed resistance inserted in the amplifier circuit. The range of frequency compensated depends on the compensation network and the mean speed of flow. For studies made in wind tunnels the lower limit of frequency is about 1 to 2 cycles per second, and the higher limit about 5000 cycles per second. The turbulent component at right angles to the stream direction, $\sqrt{v^2}$, is determined by a method based on heat diffusion [17], [18]. The characteristic length or scale of turbulence commonly used is that defined by Taylor [18] in terms of the correlation between the velocity fluctuations at two points taken on a line transverse to the stream. The correlation coefficient, R_y, is $\overline{u_1 u_2}/\sqrt{u_1^2}\sqrt{u_2^2}$, where u_1 and u_2 are the velocity fluctuations at the two points and the bars denote average values. The Taylor scale, L, is given by $\int_0^\infty R_y\,dy$. The value of the mean product $u_1 u_2$ can be measured on an electro-dynamometer, when the amplified output from the wire at point 1 is passed through the fixed coil, and that from the wire at point 2 through the moving coil [18], [19].

Grids having a uniform square mesh, constructed of circular rods or flat strips held rigidly in a stiff frame, are commonly used to create turbulence of known intensity and scale in a wind-tunnel stream. The turbulence produced by such a grid depends on the turbulence initially existing upstream of the grid, and satisfactory comparisons between theory and experiment are only possible when a grid is placed in a stream of low turbulence.

Results taken with hot wires indicate that the turbulence is isotropic in the region where the wind shadows of the rods have disappeared, and then the initial decay in intensity closely follows the law $\frac{U}{\sqrt{u^2}} = \frac{5x}{AM} + \text{const.}$ predicted by Taylor [18], where U is the stream velocity, x is the distance behind the grid, M is the grid mesh, and A is a constant for grids of the same shape. An average value of A, obtained from hot-wire results [18]–[20], is 2·1. The rate of decay departs from this law and becomes slower at and above a value of x of about 60 M. The scale of turbulence, L, increases with the distance behind a grid. A relation obtained from hot-wire results is $L/M = a + bx/M$, where a and b are constants. Average values of a and b, obtained at the National Bureau of Standards [20] from experiments on square-mesh grids, mesh sizes 1/4, 1/2, 1, 3½, and 5 inches, are 0·1467 and 0·0025 respectively.

The effects of turbulence and velocity of the stream on transition are best illustrated by considering the case of a smooth flat plate, aligned with its plane in the stream direction, the pressure gradient in

the direction of flow being zero. An increase in intensity or a decrease in scale of turbulence causes an earlier transition, that is, one occurring closer to the leading edge at a given speed, or at a lower speed for a given position on the plate. An increase in velocity of the main stream also causes an earlier transition. The position of transition in streams of low turbulence is sensitive to the accuracy of alignment of the plate, very small changes in incidence producing large changes in the position of transition. The highest value of the Reynolds number of transition, Ux/ν—where U is the stream velocity, x is the distance downstream from the leading edge, and ν is the kinematic viscosity of the fluid—so far measured on a flat plate in a wind-tunnel stream having a zero pressure gradient downstream is $23{\cdot}5\times10^5$. This value was obtained in a low-turbulence wind tunnel at Cambridge [3] ($\sqrt{u^2}/U=0{\cdot}0018$). It was not certain whether the turbulence in the tunnel was entirely responsible for fixing this upper limit, for there was evidence of an effect depending on the stream velocity and the shape of the leading edge.

The exact mechanism by which turbulence influences transition is not known, but it is believed that turbulence in the main stream outside a boundary layer superposes pressure gradients on the flow inside the layer, and as the thickness of the layer increases this eventually leads to a separation of flow. According to theory [21] the distribution of velocity in a laminar boundary layer depends on the local values of the downstream pressure gradient, the thickness of the layer, and the speed outside the layer: or, more precisely, on the value of the non-dimensional quantity $-\left(\frac{\delta^2}{\rho\nu U}\cdot\frac{dp}{dx}\right)$, denoted by Λ, where δ is the boundary-layer thickness, U is the velocity just outside the layer, ρ and ν are the density and kinematic viscosity of the fluid, and dp/dx is the downstream pressure gradient. On the assumption that the theory for steady motion holds when the pressure gradient is variable, the instantaneous velocity in the layer would depend on $(\Lambda+\Lambda')$, that is, on $-\frac{\delta^2}{\rho\nu U}\frac{d}{dx}(p+p')$, where Λ and p are steady values, and Λ' and p' are values associated with turbulence. For a thin flat plate aligned with its plane in the stream direction and with no pressure gradient $dp/dx=0$, $\Lambda=0$, $\delta^2/\nu=34\,x/U$, where x is the downstream distance from the leading edge. We therefore have

$$\Lambda'=-\frac{34x}{\rho U^2}\cdot\frac{\partial p'}{\partial x},$$

where Λ' is a fluctuating quantity. For isotropic turbulence behind a square grid, a relation for the root-mean-square of $\partial p'/\partial x$ has been obtained by Taylor [22] in terms of the intensity of turbulence $\sqrt{u^2}$ and the grid mesh M; and a relation $\frac{\sqrt{u^2}}{U}\left(\frac{x}{M}\right)^{1/5}=f\left(\frac{Ux}{\nu}\right)$ is then derived on the assumption that the root-mean-square value of $\partial p'/\partial x$ in the external stream at transition is a function of the Reynolds number at transition, Ux/ν. This relation is supported

by the results of tests made at Cambridge [3] on two flat plates, one having a sharp and the other an elliptic leading edge, mounted behind grids of mesh $\frac{1}{2}$ inch and 1 inch. The two sets of results were found to lie on the same curve, and this evidence that nose shape has no marked effect lends support to the suggestion that the influence of turbulence on transition, in the absence of pressure gradient, is confined to a local effect at the instantaneous transition point. The region of transition was observed to be fairly extensive when the fractional turbulence in the free stream was comparatively large, and the results for the mean position of transition were taken. A range of $\sqrt{u^2}/U$ from 0·01 to 0·02 was covered in these experiments, the corresponding range of the Reynolds number of transition, Ux/ν, being 7×10^5 to 2×10^5. The value of $\sqrt{u^2}/U$ in the tunnel stream without a grid was 0·0018, and, as previously mentioned, the Reynolds number of transition on a flat plate in this stream was $23{\cdot}5\times10^5$.

A sphere is commonly used as an indicator of turbulence in a stream, for there is a range of the Reynolds number UD/ν, where D is the diameter of the sphere, over which the drag is very sensitive to the turbulence in the stream. Over this range, the drag coefficient C_D, obtained by dividing the drag by the product of $\pi D^2/4$ and $\frac{1}{2}\rho U^2$, falls from about 0·5 to 0·1, and the Reynolds number at which the drag coefficient has the value 0·3 is taken as the sphere turbulence number of the stream. At this value of the drag coefficient, transition begins at a definite angular distance from the forward stagnation point, and the non-dimensional distribution of normal pressure is independent, within the accuracy of measurement, of the value of the Reynolds number [23]. By virtue of these conditions Taylor [22] has shown, by arguments similar to those used for a flat plate, that the critical Reynolds number of a sphere in an isotropic turbulent stream should be a function of $\frac{\sqrt{u^2}}{U}\left(\frac{D}{L}\right)^{1/5}$, where L is the scale of turbulence. This relation is in accordance with the results of measurements made for spheres in isotropic turbulent streams behind grids [20], [22].

It has been established that the sphere turbulence number for the atmosphere is the same as that for a non-turbulent stream in a wind tunnel. This does not mean that there are no velocity disturbances in the atmosphere (indeed it is well known that anemometer records taken in a natural wind show continual changes in speed and direction), but that the disturbances present do not affect transition in the boundary layer of a sphere. The mean speed estimated from anemometer records depends on the time interval taken in the estimation, and so the turbulence estimated also depends on the time interval, both the scale and intensity tending to increase with time. For this reason the effect of atmospheric disturbances on transition in a boundary layer depends on the nature of the problem under consideration. For the case of transition on an aeroplane wing, the time the aeroplane takes to traverse a distance of a few chords can be taken as the interval for the estimation

of mean speed of the wind. This time is of the order of 0·1 sec., so that fluctuations whose frequency are less than about 10 per second are experienced as a change in mean speed and they do not affect transition. Further, turbulences whose scale is greater than the chord length would not influence boundary-layer flow. It is for these reasons that velocity disturbances in the atmosphere do not affect boundary-layer flow, and the atmosphere behaves as if it were non-turbulent.

§ 3. EFFECT OF PRESSURE GRADIENT ON TRANSITION

When the normal pressure on the surface of a stream-line body falls in the direction of flow, the work done by the pressure gives energy to the laminar boundary layer, and this energy is used to overcome the frictional drag of the surface. When the pressure rises, energy is taken from the layer in addition to that taken to overcome the frictional drag, with the result that the fluid near the surface is soon brought to rest and the layer separates from the surface. The flow in the separated layer eventually breaks down if the Reynolds number is sufficiently high, and the fluid rejoins the surface as a turbulent layer. The position of separation for a known pressure gradient can be closely predicted from a solution of the boundary-layer equations for laminar flow [27], and experiments on smooth, polished, curved surfaces in a steady stream show that transition occurs near the predicted position of separation of the laminar boundary layer. This position occurs beyond the point of minimum pressure and in the region where the pressure is rising, and to ensure that laminar flow may extend over a large part of a curved surface the shape must be one for which the minimum pressure occurs well beyond the leading edge, the subsequent pressure rise being gradual. On aeroplane wings having a smooth surface it is possible to obtain a laminar boundary layer over at least one-third of the surface (excluding the part covered by the airscrew slipstream) at a Reynolds number as high as 8×10^6. Transition on aeroplane wings can be located with good precision in flight, and the measurements are generally made for level flight over a range of forward speeds. An increase in the speed of level flight and in the Reynolds number of the wing is then associated with a decrease in the incidence and in the lift coefficient of the wing. The increase in Reynolds number causes transition to occur earlier, whereas the change in the pressure distribution associated with a reduction of lift coefficient tends to make transition (upper surface) occur later. General experience shows that the transition point on the upper surface moves downstream, and that on the under surface upstream, with a fall in lift coefficient. The distance of the position of transition, measured along the surface of a wing from its leading edge, is found to lie on a well-defined curve when plotted against lift coefficient, and results taken on the same wing but on different days show a remarkable consistency amongst themselves. This consistency supports the opinion that transition on aeroplane wings is not affected by velocity disturbances in the atmosphere, for these disturbances vary from day to day, and from place to place.

§4. INFLUENCE OF ROUGHNESS AND SURFACE CURVATURE ON TRANSITION

Surface roughness and waviness, often when barely perceptible to the touch, cause an early transition : and this forward displacement of the position of transition increases the surface drag, in addition to any direct increase due to the departure from smoothness itself. It has been found that unevenness of surface due to a piece of tinfoil 0·002 in. thick stuck on a surface with its length at right angles to the direction of flow causes an appreciably earlier transition : and experiments in flight have shown that the drag of a smooth aeroplane wing may be appreciably increased after the aeroplane has flown through a cloud, and this increase is considered to arise from a forward movement of transition caused by mist drops deposited on the surface.

Experiments on the effect of surface curvature on transition, the pressure gradient in the direction of flow being negligible, have shown that transition occurs earlier on a concave than on a convex surface [26]. The effect is such that for the convex curvature common on the upper surface of aeroplane wings, the Reynolds number of transition is about double that measured on a smooth flat plate having the same pressure gradient down its length.

§5. NON-DIMENSIONAL TRANSITION NUMBERS

Attempts have been made to determine whether certain non-dimensional numbers formed from quantities defining the local conditions of a boundary layer have definite values at transition. The quantities taken are the thickness, δ, the velocity just outside the layer, U, the kinematic viscosity, ν, and the radius of curvature of the surface, r. The non-dimensional numbers formed from these quantities are $R_\delta = \frac{\delta U}{\nu}, \frac{\delta}{r}$, and $\lambda = \frac{\delta^2}{\nu}\frac{dU}{ds}$, where s is the distance measured along the surface from the forward stagnation point. The thickness δ can be replaced without loss of generality by the displacement thickness δ^* given by $\int_0^\infty (1-u/U)\,dy$, where u is the velocity at a distance y from the surface, or by the momentum thickness ϑ given by $\int^\infty \frac{u}{U}\left(1-\frac{u}{U}\right)dy$, since both δ^*/δ and ϑ/δ are functions of λ. The number $\frac{1}{3}\left(\frac{\delta^* U}{\nu}\right)^2$, denoted by N, has also been used. For flow along a flat plate, without a downstream pressure gradient, Pohlhausen's method of solution of the boundary-layer equations for laminar flow [21] gives $R_x = \left(\frac{\delta^* U}{\nu}\right)^2 / 3{\cdot}06$, where x is the distance from the leading edge of the plate. The number N is therefore approximately equal to the x-Reynolds number of a flat plate associated with an observed value of R_{δ^*}, and gives the stage of development of the laminar layer at transition.

If transition is solely dependent (in the absence of turbulence in the free

stream) on N, λ, and δ^*/r, observed values of N at transition plotted against the corresponding values of λ, for constant values of λ^*/r, should lie on a family of curves. Flight experiments show that the values of λ at the position of transition on aeroplane wings lie between 0 and -7, and that within this range there is no apparent relation between N and λ at a constant value of δ^*/r, or between N and δ^*/r at a constant value of λ. It would, therefore, appear that time-average values of the quantities δ^*, U, and dU/ds do not take into account all the factors which influence the flow in a boundary layer up to the position of transition. Velocity fluctuations in the boundary layer, such as those arising from very slight waviness or unevenness of the surface or associated in some way with vibration or swaying of the aeroplane, were not recorded in the flight experiments, and if such fluctuations were present they must be regarded as a possible contributory cause of transition. It is known that long-period fluctuations are present in a laminar layer, and calculations [24] have shown that such fluctuations become greatly magnified as the surface is approached. For this reason it is to be expected that such disturbances may be powerful agents in the breakdown of laminar flow. Further research on the possibility of specifying transition in terms of non-dimensional numbers is needed, and is now in progress. Present indications are that no simple specification is likely when turbulence is present in the free stream, in addition to an adverse pressure gradient [25].

REFERENCES

(1) Melvill Jones. *J. Aero. Sci.* **5**, no. 3 (1938).
(2) Stephens and Haslam. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1800 (1938).
(3) Hall and Hislop. *Rep. Memor. Aero. Res. Comm., Lond.* no. 1843 (1938).
(4) Fage and Falkner. *Proc. Roy. Soc.* A, **129** (1930).
(5) Hansen. *Z. angew. Math. Mech.* **8** (1928).
(6) van der Hegge Zijnen. *Proc. K. Akad. Wet. Amst.* (Thesis, Delft, 1924).
(7) Dryden. *J. Aero. Sci.* **6**, no. 3 (1939).
(8) Schubauer. *Proc. Fifth Intern. Congress Appl. Mech.* (1938).
(9) Dryden. *Rep. Amer. Nat. Adv. Comm. Aero.* no. 562 (1936).
(10) Dryden. *J. Washington Acad. Sci.* **25**, no. 3 (1935).
(11) Ziegler. *Proc. K. Akad. Wet. Amst.* **35**, no. 3 (1932).
(12) Simmons and Brown. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1547 (1934).
(13) Simmons and Dewey. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1335 (1930).
(14) Burgers. *Proc. K. Akad. Wet. Amst.* **29** (1926).
(15) Mock and Dryden. *Rep. Amer. Nat. Adv. Comm. Aero.* no. 448 (1933).
(16) Dryden and Kuethe. *Rep. Amer. Nat. Adv. Comm. Aero.* no. 320 (1929).
(17) Schubauer. *Rep. Amer. Nat. Adv. Comm. Aero.* no. 524 (1935).
(18) Taylor *Proc. Roy. Soc.* A, **151** (1935).
(19) Hall. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1842 (1938).
(20) Dryden, Schubauer, Mock and Skramstad. *Rep. Amer. Nat. Adv. Comm. Aero.* no. 581 (1937).
(21) Howarth. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1632 (1934).
(22) Taylor. *Proc. Roy. Soc.* A, **156**, 888 (1936).
(23) Fage. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1766 (1936).
(24) Dryden. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 562 (1936).
(25) Fage. *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1765 (1936).
(26) Clauser and Clauser. *Tech. Note Amer. Adv. Comm. Aero.* no. 613 (1937).
(27) Howarth. *Proc. Roy. Soc.* A, **164** (1938).

THE PROPERTIES OF LIQUID HELIUM

By H. JONES, Ph.D.
Imperial College, London

§ 1. INTRODUCTION

During the last three years many investigations of the properties of liquid helium have been made in the Mond Laboratory in Cambridge, at Leiden, Oxford and elsewhere. These investigations have dealt almost exclusively with the transport phenomena of the low-temperature modification of the liquid. New and unexpected effects which are peculiar to liquid helium have been brought to light and, as is natural, along with the experimental discoveries a good deal of interesting theoretical speculation has been put forward from time to time. This report will be mainly devoted to describing these newly investigated transport properties and the attempts which are being made to formulate the results. In these attempts great difficulties are experienced. For instance, an inspection of the data published by different workers on the flow properties of liquid helium, i.e. data of experiments designed to determine the viscosity, reveals widely divergent results and leads one to suppose that the situation is almost chaotic. Recent work by Allen and Misener [1], however, has done much to clear up the position in this respect, and a discussion of their results in § 4 (*c*) shows that in fact no discrepancies in order of magnitude exist.

The work which followed the discovery by Rollin [2] and Keesom [3] in 1935 of the exceptionally high heat conductivity of liquid helium led to the discovery, in the Mond Laboratory [4], of the peculiar thermomechanical effect which is described in § 4 (*a*). Certain thermodynamical considerations concerning this effect recently put forward by H. London [5] are given in § 4 (*e*). Although at present there is no detailed theory of the constitution of liquid helium in its low-temperature modification, there is reason to hope that, as a result of a closer study of the thermomechanical effects, a satisfactory phenomenological theory may soon emerge. In this connection the investigations of Daunt and Mendelssohn [6] on the transfer of liquid by mobile surface films seem likely to prove of basic importance.

In this report only a brief account is given of the equilibrium properties of liquid helium, which have now been well established for many years. Data of this nature, accumulated mainly by the researches of Keesom [7] and his collaborators, form a basis for the more recent investigations. For example, the accurate determination of the vapour-pressure : temperature relation forms the basis for all temperature measurements in the range from 1 to 4 degrees absolute.

§ 2. PROPERTIES OF LIQUID HELIUM IN THERMAL EQUILIBRIUM

Helium was liquefied for the first time in 1908 by Kamerlingh Onnes. It boils under atmospheric pressure at 4·22° K. and has a critical temperature at 5·2° K. The density is approximately 0·147 gm. per cm^3 At 2·19° K. under its own vapour pressure liquid helium undergoes a remarkable transformation. The high-temperature modification of the liquid which exists between 2·19° K. and 4·22° K. is known as He I, and this liquid cannot be said to present any very unusual features. The low-temperature modification which exists from absolute zero up to 2·19° K. under its own vapour pressure is known as He II. The critical temperature and pressure at which the transformation occurs define the λ-point. We give below a few values of the temperature and pressure and the corresponding density at the λ-point :—

T (°K.)	2·190	2·125	1·990	1·915	1·835	1·753
Pressure (atm.) ..	0·025	4·4	13·5	18·7	24	29·91
Density (gm./cm). ..	0·1473	0·1541	0·1666	0·1714	0·1757	0·1796

The specific heat of He II increases very rapidly (approximately as T^5) from temperatures of about 1° K. to the λ-point, at which there is a discontinuous fall in its value. Specific heats which depend on temperature in this way are well known in other branches of physics. For example, the order-disorder transformation in alloys gives rise to such a specific-heat : temperature relation. Other examples are the Curie point transformation of a ferromagnetic, and the transformation from the supraconducting to the non-supraconducting state of a metal. Some values of the specific heat of liquid helium, determined by W. H. Keesom and A. P. Keesom [7], are given below :—

T (°K.)	1·224	1·231	1·314	1·487	2·041	2·184	2·188
$C_{vap.}$ (cal. per gm. per degree) ..	0·0923	0·1001	0·1404	0·278	1·34	2·88	5·83

$C_{vap.}$ denotes the value of the specific heat of the liquid under its saturated vapour pressure. W. H. Keesom and A. P. Keesom have also given values of the specific heat at constant volume. The results for a density 0·1683 gm. per cm^3 are given below

T (°K.)	1·211	1·391	1·728	1·910	1·959	1·965	2·006
C_v (cal. per gm. per degree)	0·153	0·319	0·843	1·393	1·619	1·949	0·503

The determination of the vapour pressure of liquid helium for different temperatures down to one degree absolute has been made at Leiden. The results are expressed by empirical formulae giving the pressure in terms of temperature. The temperature scale thus established in 1932 has been corrected by Schmidt and Keesom [8] in 1937 ; the corrections amount to 1 or 2 per cent.

It has been observed that liquid helium will not solidify under its own vapour pressure at the lowest temperature obtainable. It will, however, solidify under pressure. For instance the solid is formed at 1° K. and a pressure of 25 atm. For higher temperatures, higher pressures are required, e.g.

29·9 atm. are necessary at 1·75° K. In the neighbourhood of 1° K., however, the required pressure appears to remain about 25 atm., and it is almost certain that this would be approximately the pressure required to produce the solid at absolute zero. Since the pressure required to effect solidification in the neighbourhood of 1° K. and below is almost independent of the temperature, it follows, from the well-known equation of Clapeyron, that there is no change in entropy on passing from liquid to solid helium at these very low temperatures. This is in accordance with the expectation that both liquid and solid should have zero entropy at the absolute zero of temperature. Below are given approximate values of the entropy of He II in cal. per gm. per degree for a few different temperatures taken from the curves published by Keesom [7]:—

T (°K.)	1·5	1·8	2·0	2·15
Entropy (cal. per gm. per degree)..	0·06	0·15	0·25	0·35

It is useful to consider the heat energy in He II as made up of two parts in which the energy takes different and distinguishable forms. This is also the case, for example, in β-brass near the critical temperature where the heat energy is partly energy of elastic vibrations and partly energy of disorder. In He II the heat energy associated with elastic vibrations appears to be very small. This conclusion is drawn from the $C_v : T$ curve. We should expect the specific heat arising from elastic vibrations to vary as T^3, but since the observed specific heat, from about 1° K. to the λ-point, varies approximately as T^5, we are led to conclude that the elastic vibrations contribute only a small part to the total heat content. The greater part arises in some process which is not yet understood, and which may, therefore, be called the anomalous heat. Thus if, for theoretical reasons, we wished to know the anomalous specific heat or entropy, corrections would have to be applied to the values given above by subtracting the contribution arising from the elastic waves. Recently, by measurements made in the region of temperature obtained by adiabatic demagnetization, Simon has obtained values for the specific heat which vary as T^3 and which, therefore, presumably represent the vibrational part.

We have seen that the entropy of He II vanishes at absolute zero, and as entropy is a measure of the logarithm of the number of ways in which a state can be realized, it follows that the state of liquid helium at absolute zero can be realized in only one way. This has led several investigators to suppose that at low temperatures liquid helium has some kind of space-ordered structure over regions containing a moderately large number of atoms. An experimental investigation of this point has been carried out by Keesom and Taconis[9]. They examined the reflections of x rays from a column of liquid helium at 1·6° K. as well as above the λ-point. No essential difference between the reflections in the two cases could be observed. The x-ray work is exceedingly difficult on account of the minute scattering power of the helium atoms and the trouble experienced in making long exposures with liquid helium. While the experiments give no hint of a space order in He II, they cannot be taken as ruling out the possibility of a limited amount of order at lower temperatures.

§3. DISCUSSION OF THE EQUILIBRIUM PROPERTIES

When we consider the binding energy and density of liquid helium at low temperatures an interesting fact at once emerges. In materials of common experience, e.g. NaCl, the binding energy is very nearly equal to the negative potential energy arising from the attraction between the constituent atoms; thus the atomic volume is determined by the volume per atom for which this potential energy is a minimum. This is not exactly true, however, because the vibrations associated with the thermal energy expand the solid slightly beyond the volume corresponding to the minimum potential energy. Now a well-known feature of quantum theory is that, even at absolute zero, the atoms are not at rest but possess a certain zero-point kinetic energy. This zero-point energy and the consequent increase in the atomic volume beyond the value for minimum potential energy are in ordinary substances exceedingly small compared with the binding energy and the atomic volume themselves.

The zero-point energy of liquid He II, which has been deduced from thermal measurements, is of the order of 70 cal. per gm. atom. The total binding energy, on the other hand, is only 14 cal. per gm. atom. It thus appears, as F. London [10] has shown by more detailed calculations, that the zero-point energy is almost as large, in liquid helium, as the negative potential energy, and that this relatively large value leads to the high atomic volume to which ultimately many of the peculiar properties of He II seem largely to be due. If, for example, the expansive effect of the zero-point energy is balanced by external compression we obtain the solid helium which behaves normally, having, for instance, the usual T^3 law of specific heat.

Another effect of the high zero-point energy is to make the most stable state of a collection of helium atoms one of low co-ordination number.* It appears probable, from this consideration, that the average number of nearest neighbours to an atom in liquid helium is not much greater than 4.

To explain the λ-point phenomenon two hypotheses have been put forward. Fröhlich has suggested that the effect might be analogous to the order-disorder process which is known to occur in alloys, e.g. β-brass. This hypothesis supposes that at low temperatures liquid helium has a large degree of space order with low co-ordination number, and that as the temperature rises an increasing number of atoms move away from their ordered positions. At the critical temperature almost the whole order has disappeared. In the known transformations in which the specific heat jumps discontinuously, but in which there is no discontinuity in the entropy, the mechanism is generally of this type where there is a change from an ordered to a disordered state. In the case of liquid helium alone there appears to be another possibility.

It was shown by Einstein [12] in 1925 that a perfect gas satisfying the Einstein-Bose statistics (to which helium conforms) undergoes a discontinuous change at low temperatures. For instance, below a certain temperature, the specific heat increases as $T^{3/2}$, while beyond this temperature it decreases slightly so that the specific-heat : temperature curve has a sharp kink at the critical

* Cf. the calculations of F. London [10].

point. This transformation differs from a λ-point in which the specific heat itself is discontinuous. Einstein gave the following formula for the critical temperature :

$$n\left(\frac{h^2}{2\pi m k T}\right)^{3/2} = 2{\cdot}612, \qquad \ldots.(1)$$

where n is the number of particles of mass m per unit volume, h Planck's constant, and k Boltzmann's constant.

F. London [13] has pointed out the interesting fact that if the values appropriate to liquid helium are substituted in the above expression the critical temperature is found to be 3·09° K., which is not far from the λ-point. Whether the λ-point phenomenon is due to the quantum statistics which the atoms satisfy or is simply controlled by the nature of the interatomic forces, as in other known cases of this type, is still an open question.

There is one conclusion which may, perhaps, be drawn with not too great uncertainty. This is that in liquid helium the anomalous heat energy is not distributed uniformly amongst all the atoms of the liquid but is contained in a certain number of excited atoms which a study of the transport effects suggests may have a limited free translational motion inside the liquid rather like the electrons in a metal.* If the liquid at the lowest temperature has an open ordered structure the excited atoms may be pictured as moving through the interstices of a loosely connected lattice.

§ 4. THE TRANSPORT PROPERTIES OF HELIUM

There are four main transport effects, viz. (1) The thermomechanical effect, generally referred to as the fountain effect. (2) The transfer of liquid helium over surfaces as mobile films. (3) Viscous flow. (4) Heat conductivity. These are dealt with in this order since the first two are peculiar to He II and almost certainly play some part in the experimental arrangements designed to measure the viscosity and the heat conductivity.

(a) *The thermomechanical effect.*

The simplest apparatus which will show this effect is indicated in the figure opposite. A tube of the form shown is partly immersed in He II. The lower end forms a capillary ; both ends are open, and arrangement is made to heat the liquid in the upper part. A steady flow of heat is thereby maintained down the capillary. Under these conditions it is observed that the liquid inside the tube rises above the level outside, showing the existence of a pressure in the reverse direction to the heat flow. A much more pronounced effect is observed when the capillary is replaced by a wider tube filled with emery powder. Allen and Reekie [14] have made measurements with such an arrangement. The temperature of the bath was determined by means of the vapour pressure, whilst the temperature inside the tube when heat was applied was measured by a sensitive resistance thermometer. The difference in temperature between the liquid in the upper part of the tube and

* Cf. Michels, Bijl and de Boer [28], and also Jones [27].

that in the bath was generally of the order of 10^{-3} degree. We shall denote this temperature difference by ΔT. Measurements were made of the difference in level for various ΔT and mean temperatures. Level differences are converted into pressure heads in dynes per cm^2, denoted by Δp, by multiplying

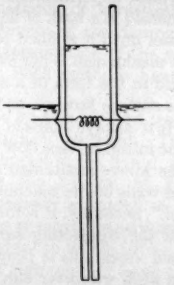


Figure 1. Apparatus to show the thermomechanical effect.

by the density and g. Some results of these experiments are shown in figure 2. For details of the apparatus, which is only schematically indicated here, reference should be made to the original paper.

A striking feature of these results is that the reaction pressure Δp does not

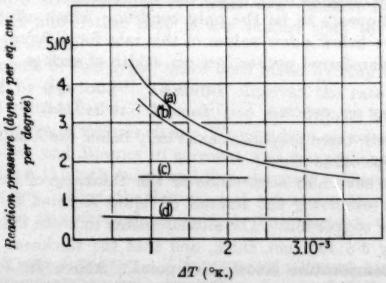


Figure 2. The variation of the reaction pressure with temperature difference for different mean temperatures. (*a*) $T = 1{\cdot}789°$ K.; (*b*) $T = 1{\cdot}930°$ K.; (*c*) $T = 1{\cdot}393°$ K.; (*d*) $T = 1{\cdot}241°$ K.

depend linearly on ΔT. In fact it appears doubtful whether a limiting value of $\Delta p/\Delta T$ exists in the limit $\Delta T \to 0$. The curve shown in figure 2 for $T = 1{\cdot}241°$ K. appears parallel to the ΔT axis, but there is reason to believe that it also would rise steeply for small temperature differences. We return to a discussion of this effect in (*e*) below.

(b) *Surface transfer of He II*

Daunt and Mendelssohn [6] have performed a series of experiments which have elucidated the phenomena associated with the transfer of liquid He II over surfaces in the form of mobile films. A simple idealized arrangement which gives an idea of a typical experiment is as follows. An empty vessel open at the top is partly immersed in a bath of liquid He II. It is then observed that He II collects in the vessel until it is filled to the bath level. It has been abundantly verified that the mechanism is not evaporation and recondensation but is, in fact, a flow of liquid in the form of a mobile surface film.

The characteristics of the surface flow are as follows. First, the flow takes place in such a way that He II always collects at the lowest available level. This is the case even when the initial surface flow takes place from low to higher points as, for example, in the above-mentioned arrangement, where the films must first flow up the outside walls before reaching the lower level on the inside of the partly immersed vessel. Secondly, it is found that the flow is accurately proportional to the width of the connecting surface. For instance, the rate at which the above-mentioned vessel fills is proportional to the circumference of the narrowest part of the neck whatever the shape of the remaining parts of the vessel may be. It is thus possible to define and measure a rate of transfer in cubic cm. per sec. per cm. width of connecting surface. Thirdly, the rate of transfer so defined is found to be practically independent of the difference of height between the levels of the liquid. For instance, as above, the vessel fills at the same rate from the start until the levels coincide ; the transfer then stops. Also, as far as can be judged, the rate of transfer is independent of the material over which the films move.

Temperature appears to be the only condition which affects the rate of transfer. We give below a few values of this rate for different temperatures. V is the volume transferred per sec. per cm. width of surface.

T (°K.)	1·0	1·5	1·7	1·9	2·1	2·19
$V.10^5$ (c.c. per sec. per cm.) ..	7·4	7·5	6·3	4·0	1·2	0

Mobile films with these properties exist only below the λ-point ; in He I on transfer of this type takes place.

Measurements have also been made of the thickness of the He II films.* This was done by observing the amount of liquid required to cover the dry surface of a roll of copper foil. The measurements indicate that films of He II are approximately $3{\cdot}5.10^{-6}$ cm. thick, and that the thickness does not vary greatly with the temperature below the λ-point. Above the λ-point the He I films do not appear to be more than 10^{-7} cm. thick. The average velocity through a He II film below 1·5° K. is therefore approximately 21 cm. per sec.

The mechanism by which surface transfer takes place is not at present understood. The fact that flow is independent of level difference is, in particular, a great difficulty. A possibility is that the surface flow consists of a flow of unexcited atoms, i.e. of the liquid without its anomalous heat, and for this there is some evidence. In this case heat would be evolved at the point where the

* An estimate of the thickness has also been made by Kikoin and Lasarew [15].

surface flow begins and would be absorbed at the place where the flow ends. The rate of flow may thus be controlled principally by the rate at which the heating and cooling produced could be balanced in the reservoirs.

(c) *Viscosity and flow properties*

There are two general methods of investigating the viscous properties of a liquid:—(*a*) The measurement of the damping of the motion of a body immersed in it, and (*b*) the determination of the rate of flow through tubes or between parallel plates, etc. Both methods have been used with He II and some curious results obtained.

(*a*) The first investigations by this method were made by Burton [16], and by Wilhelm, Misener and Clark [17] in 1935. They measured the damping of an oscillating cylinder suspended in the liquid, and obtained a value for the viscosity of He I, just above the λ-point, of $2{\cdot}7.10^{-5}$ dyne-sec. per cm^2. Just below the λ-point the viscosity appeared to be about ten times smaller. For comparison of magnitudes it may be mentioned that helium gas at 15° K. has a coefficient of viscosity equal to $2{\cdot}946.10^{-5}$ c.g.s. units.

More recently Keesom and Macwood [18] have carried out a series of experiments by measuring the damping of an oscillating disc in liquid helium. A disc is more suitable for accurate viscosity determinations than a cylinder. The accompanying table gives some of their results:—

T (° K)	1·20	1·50	1·75	2·00	2·10	2·163
η (c.g.s. units $\times 10^{-5}$) .	0·15	0·20	0·38	0·83	1·20	1·80

In these experiments care was taken to keep the velocity of the liquid small enough to ensure stream-line flow. This was checked by estimating the Reynolds number and noting that it was smaller than the critical value at which turbulence sets in.

(*b*) Experiments of this type * have been made by Kapitza [20], Allen and Misener [1], and by Burton [19]. Kapitza observed the rate of flow between parallel plates, whilst Allen and Misener, and also Burton, have examined the flow down capillaries of various sizes. We give here an account of the results obtained by Allen and Misener in a recent and comprehensive investigation of the flow of He II through tubes and fine channels. The work may be considered in two parts. First, the flow through tubes of radius 10^{-2} cm. or more, and secondly, the flow through exceedingly fine channels whose width is of order 10^{-5} cm.

Flow through wide capillaries. There are two different kinds of flow which can occur in smooth tubes, viz. stream-line flow and turbulent flow. The factors which determine which type of flow shall exist are the average velocity of the liquid in the tube, denoted by $\bar{u}$, the radius of the tube r, and the kinematic viscosity η/ρ, i.e. viscosity divided by density. The circumstances which favour the setting up of turbulence are large tubes, high velocities, and small kinematic viscosity.

* Measurements of viscosity by a flow method have also been made by Giauque, Stout and Barieau [26].

It is known from experience with liquids and gases of all kinds that when the pure number

$$R=r\bar{u}\rho/\eta \qquad \ldots.(2)$$

(known as the Reynolds number) exceeds about 1000, turbulence is likely to occur. For values of $R<1000$ the flow is stream-line and the average velocity in the tube is given by Poiseuille's formula

$$\bar{u}=\frac{r^2}{8\eta l}.\Delta p, \qquad \ldots.(3)$$

where Δp is the pressure difference at the ends of the tube and l is the length; the other symbols are as defined above.

When R is greater than about 2000 the flow is in general completely turbulent, and experiments have shown that the average velocity may then be well represented by the formula of Blasius [21],

$$\bar{u}^2=2r\Delta p/\lambda l\rho, \qquad \ldots.(4)$$

where $\lambda=0{\cdot}133R^{-1/4}$.

The average velocity $\bar{u}$ in turbulent flow is, in general, less than the value which would be obtained under the same conditions were the flow stream-line, i.e. as given by formula (3). It can thus happen, as with water for example, that $\bar{u}$ for a constant pressure head may decrease as the viscosity is decreased (as in the case of water by raising the temperature). This happens when the conditions are in the critical region between those favourable to the two kinds of flow.

When $\bar{u}$ has been determined by experiment for different values of Δp one way of showing the nature of the flow is to calculate $d\log\bar{u}/d\log\Delta p$. It will be seen from formula (3) that this quantity is 1 for stream-line flow and 4/7, according to formula (4), for turbulent flow. Intermediate values would indicate that the conditions are in the critical region between the two types of flow.

We give below data, taken from Allen and Misener's paper, of the flow of He II through a tube of radius $r=4{\cdot}38.10^{-2}$ cm. and length $l=20$ cm.; Δp is expressed in dynes per cm^2 and $\bar{u}$ in cm. per sec.

Table 1

T (°K.)	$\frac{d\log\bar{u}}{d\log\Delta p}$	$\frac{\Delta p=160}{\bar{u}}$	$\frac{\Delta p=15}{\bar{u}}$
2·163	0·80	16·3	2·5
2·006	0·63	14·4	—
1·751	0·57	13·2	—
1·191	0·54	12·2	3·5

It is seen from the second column that at 2·163° K. the conditions approach those for stream-line flow. Of the two pressure heads the smaller will be the more favourable, as it leads to the smaller average velocity. Taking the

values $\Delta p = 15$ dynes per cm^2, $\bar{u} = 2{\cdot}5$ cm. per sec. with the length and radius given, we find according to equation (3) $\eta = 7{\cdot}2.10^{-5}$ c.g.s. units. Using this value of the viscosity, the Reynolds number according to equation (2) is 228, which agrees with the assumption of stream-line flow.

At the same temperature the value of the coefficient of viscosity given by Keesom and Macwood is approximately $1{\cdot}8.10^{-5}$ c.g.s. units, which is about one quarter of the value given by the flow experiments.

The results of three experiments at the same temperature by Burton [19] give for η the values $7{\cdot}9.10^{-5}$, $7{\cdot}8.10^{-5}$, and $5{\cdot}4.10^{-5}$ c.g.s. units.

Next we examine the flow at lower temperatures and higher pressures. It is immediately obvious, here, that there is no stream-line flow, and it is therefore practically impossible from the data to calculate a viscosity with any accuracy. It is interesting, however, to see whether it is possible to account for the observed flow as the setting up of turbulence or whether one is forced to consider some additional process peculiar to He II, as is certainly the case, for instance, for flow in very narrow channels.

To do this we calculate the average velocity for different viscosities from formulae (3) and (4) for pressures 160 dynes per cm^2 and 15 dynes per cm^2 and the dimensions of the tube as already stated. The results are given below.

Table 2

η (c.g.s. units)	$\Delta p = 160$ dynes/cm^2		$\Delta p = 15$ dynes/cm^2	
	$\bar{u}$ st. line	$\bar{u}$ turb.	$\bar{u}$ st. line	$\bar{u}$ turb.
$7{\cdot}2.10^{-5}$	27·0	14·6	2·5	3·76
$3{\cdot}3$,,	58·1	16·4	—	—
$1{\cdot}5$,,	126	18·2	—	—
$0{\cdot}6$,,	320	20·7	30·0	5·36

The average velocity of such a fluid would be, therefore, 2·5 cm. per sec. at 15 dynes per cm^2, when $\eta = 7{\cdot}2.10^{-5}$, but when $\eta = 0{\cdot}6.10^{-5}$ the velocity would be 5·36 cm. per sec., for at the lower value of η the stream-line flow gives place to turbulent flow. Similarly at 160 dynes per cm^2 $\bar{u}$ would start, for high η, somewhere near 27 and decrease to 20·7 for $\eta = 0{\cdot}6.10^{-5}$.

To make comparison with He II we require a knowledge of the temperature-dependence of the viscosity. In the first column of table 2 the value of $\eta = 7{\cdot}2.10^{-5}$ c.g.s. units is that which was deduced from the flow experiment at $\Delta p = 15$ dynes per cm^2 and $T = 2{\cdot}16°$ K. assuming stream-line flow, and consequently the observed value of $\bar{u} = 2{\cdot}5$ cm. per sec. appears in the fourth column. As we have said, this value of η is approximately four times that given by Keesom and Macwood for this temperature. The remaining three values of η in column 1 are also approximately four times the values given by Keesom and Macwood for the temperatures 2·00° K., 1·75° K. and 1·20° K.

respectively. The values of η in table 2 are the ones we shall assume to correspond to the temperatures in table 1; they form a plausible compromise between the available data.

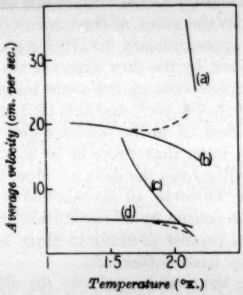


Figure 3. The variation of the average velocity with temperature for streamline flow ((*a*) and (*c*)) and turbulent flow ((*b*) and (*d*)).

Figure 3 shows the values of $\bar{u}$ from table 2 plotted against T with the above η, T correlation. The dotted lines are sketched in to show what may be expected in the transition from stream line to turbulent flow. It will be seen that the lines reproduce roughly the behaviour of He II. The velocities are not far out in magnitude and the temperature variation for the two pressures is correctly reproduced (cf. table 1). It seems permissible, therefore, to draw two conclusions. First, that so far as flow in wide capillaries is concerned, no new process is involved to any appreciable extent, and secondly, that a viscosity varying with temperature as found by Keesom and Macwood is not incompatible with the flow experiments.

Allen and Misener give data for the average velocity in tubes of different radii where the length of the tubes, the temperature and the pressure head remain the same. The data for the relatively wide capillaries are summarized in table 3, and are plotted in figure 4. The length of each tube was 20 cm.

Table 3

r (cm. $\times 10^{-3}$)	$T = 1{\cdot}20°$ K.		$T = 2{\cdot}16°$ K.	
	$\Delta p = 160$ $\bar{u}$ cm./sec.	$\Delta p = 15$ $\bar{u}$ cm./sec.	$\Delta p = 160$ $\bar{u}$ cm./sec.	$\Delta p = 15$ $\bar{u}$ cm./sec.
0·5	5·7	2·9	1·4	0·42
1·53	7·3	2·8	6·0	1·1
4·38	12·2	3·5	16·1	2·5

Table 4 and figure 4 (dotted lines) show the average velocities calculated according to formulae (3) and (4). At the lower temperature the Blasius formula (4) for turbulent flow has been used and at the higher temperature the Poiseuille formula (3) for stream-line flow. The value of the density

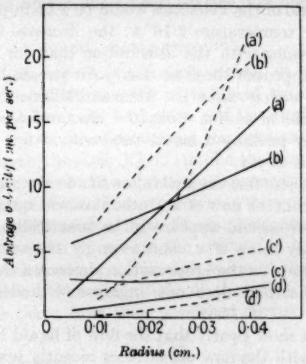


Figure 4. The variation of the average velocity with radius.

(*a*) $\Delta p = 160$ dynes per cm^2. $T = 2{\cdot}16°$ K. (*c*) $\Delta p = 15$ dynes per cm^2. $T = 1{\cdot}2°$ K
(*b*) $\Delta p = 160$ „ „ „ $T = 1{\cdot}2°$ K. (*d*) $\Delta p = 15$ „ „ „ $T = 2{\cdot}16°$ K.
(*a'*), (*b'*), (*c'*), (*d'*) are calculated curves corresponding to (*a*), (*b*), (*c*), (*d*) respectively.

was taken as 0·147 gm. per cm^3. To obtain these figures we have, as before, taken the viscosity consistently four times higher than the values of Keesom and Macwood. Although the differences are considerable, they hardly seem great enough, in view of the difficulty of the experiments, to justify the assumption that anything other than ordinary hydrodynamic flow occurs to any appreciable extent in tubes of radius greater than 10^{-2} cm.

Table 4

r (cm. $\times 10^{-2}$)	$T = 1{\cdot}2°$ K. $\eta = 0{\cdot}6.10^{-4}$		$T = 2{\cdot}16°$ K. $\eta = 7{\cdot}2.10^{-4}$	
	$\Delta p = 160$ $\bar{u}$	$\Delta p = 15$ $\bar{u}$	$\Delta p = 160$ $\bar{u}$	$\Delta p = 15$ $\bar{u}$
1	7·3	1·88	1·39	0·130
2	11·9	3·10	5·56	0·521
3	16·0	4·14	12·50	1·172
4	19·6	5·08	22·22	2·083

The variation of average velocity with length for constant radius, pressure head, and temperature has also been given by Allen and Misener. Again rough agreement can be obtained. For instance, in a tube 10 cm. long at $T=1{\cdot}2°$ K. the average velocity was 10·1 cm. per sec.; in a tube 40 cm. long, 5·2 cm. per sec. In turbulent flow, according to equation (4), $\bar{u} \propto l^{-4/7}$, so that in the above case the ratio of the velocities would be 2·21 in place of the observed 1·95. At the higher temperature 2·16° K. the decrease in $\bar{u}$ with increasing length is greater, agreeing with the assumption that, for the higher temperature, the conditions approach those necessary for stream-line flow.

Flow through very narrow channels. Allen and Misener have studied the flow of He II through a tube of radius $1{\cdot}62.10^{-3}$ cm. and also through very fine channels produced by packing a metal tube with a few thousand fine wires and drawing the whole down.

The flow through these fine channels shows two very striking characteristics which distinguish it from the flow of any other known liquid. First, the volume of liquid delivered per second appears to be practically independent of the pressure head, certainly below 2° K., and secondly, the average velocity of flow in the channels increases as the cross-section decreases though the volume of liquid delivered per second does not increase with decreasing size of the channels.

These two features show clearly that the flow of liquid He II in narrow tubes is closely connected with the flow on surfaces recently investigated by Daunt and Mendelssohn and described in § 4 (*b*).

Allen and Misener show the dependence of flow upon pressure head by giving the quantity $d \log \bar{u}/d \log \Delta p$. As we have seen, this number is unity for stream-line flow and approximately 0·57 for turbulent flow. If we suppose that a fraction f of the liquid delivered through the tube is carried by surface flow or some analogous process which leads to complete pressure independence, and the remainder $(1-f)$ of the flow is turbulent, it follows at once that

$$d \log \bar{u}/d \log \Delta p = 0{\cdot}57(1-f). \qquad \ldots.(5)$$

If we had supposed that the remaining fraction $(1-f)$ was carried by stream-line flow we should have had unity in place of the fraction 0·57.

For the narrowest channels investigated by Allen and Misener, formed by a bundle of 2000 wires, $d \log \bar{u}/d \log \Delta p$ is practically zero below 1·8° K., which implies $f=1$. If the total internal surface per cm. length is multiplied by the surface-flow constant of Daunt and Mendelssohn, i.e. $7{\cdot}5.10^{-5}$ cm^2 per sec. per cm., the result is found to be close to the total volume delivered per second. This case appears, therefore, not essentially different from the surface-flow experiments.

The single capillary of radius $1{\cdot}62.10^{-3}$ cm. is an interesting intermediate case between almost pure hydrodynamical flow and surface flow. Here at low temperatures, less than 1·8° K., $d \log \bar{u}/d \log \Delta p = 0{\cdot}13$. This would give $f=0{\cdot}77$. The total flow in this tube at low temperatures, which is only slightly dependent on T and Δp, is about $3{\cdot}3.10^{-5}$ cubic cm. per sec. Thus about

$2{\cdot}5.10^{-5}$ c.c. per sec. appears to be pressure-independent flow. If, however, we multiply the internal periphery by the surface-flow constant, we obtain only $0{\cdot}75.10^{-6}$ cm. per sec. These experiments seem to show, therefore, that that part of the flow in capillaries which is independent of pressure and which becomes significant in small capillaries is still considerably greater than that which would be given by the known free surface flow down the inside walls.

(d) *Heat conductivity*

Very little detailed experimental work exists, at present, on the heat conductivity of He II. Measurements of the heat transported along a tube filled with He II have been made at Leiden and Cambridge. If ΔQ represents the heat carried per sq. cm. cross-section of the tube and ΔT the temperature difference per cm. length, the ratio $\Delta Q/\Delta T$, unlike the case in ordinary materials, depends on the magnitude at ΔT and does not appear to reach a limit for small ΔT. The experimental results form, therefore, a very complicated set of data.

Keesom and Saris [22] used a tube 35 cm. long and 0·30 mm. diameter. They express their results by giving the values of $k=\Delta Q/\Delta T$ for different temperatures and heat currents. For each value of the heat current there is a separate k, T curve. The magnitude of the effect can be seen from the following few values taken from their published curves :—

T (°K.)	1·22	1·5	1·7	1·8
k (watts/cm. degree)	1000	2000	2000	1300
Heat current (watts/cm^2) ..	0·098	0·227	0·578	1·040

For immediate comparison we note that copper at room temperature has a heat conductivity of 4 watts per cm. degree.

Experiments have also been made by Allen and Reekie [14] on the heat transported through a tube packed with fine emery powder. From these experiments it is not possible to obtain the absolute value of the heat conductivity, but its variation with ΔT for different mean temperatures is found and the rapid increases with diminishing ΔT clearly demonstrated.

Allen, Peierls and Uddin [23] have also obtained data for the heat conductivity of He II, but their published results need rather a large correction due to the thermomechanical effect, which at the time of the experiments had not been recognized.

At temperatures between 0·2° K. and 0·5° K., obtained by the adiabatic demagnetization of a paramagnetic salt, Kürti and Simon [24] have shown that the heat conductivity of He II becomes quite small. They give the following values :—

T (°K.)	0·5	0·4	0·3	0·2
$k.10^3$ (cal./sec. degree cm.) ..	2·2	1·4	0·7	0·2

In this temperature range the specific heat varies as T^3 (cf. § 2).

The variation of heat conductivity with temperature gradient has not yet been satisfactorily explained. It will be seen that it is necessary to explain an anomalously large transport of heat for a very small temperature difference.

It has been suggested that this is a result of the thermomechanical effect. The small temperature gradient sets up a circulation, in the sense that unexcited atoms flow towards the hotter part (presumably along the walls), whilst in the centre of the tube an ordinary flow of hot liquid takes place towards the lower temperature. That a circulation of this sort does in fact exist, and contribute to the heat transport, seems almost certain from the observations on the fountain effect already described. It is not clear, however, why heat transported by this method should not also be proportional to the temperature gradient. The essential difficulty therefore at present remains unsolved.

At temperatures in the neighbourhood of 1·2° K. and with high heat currents the conductivity seems to approach independence of temperature gradient. Under these conditions the value of the conductivity is still high, and only further experimental investigations can decide whether this represents a true homogeneous conductivity or is the result of a complex system of convection currents set up by the fountain effect.

(e) *Thermodynamical considerations of the fountain effect*

In liquid He II, experiments have indicated that apart from the hydrodynamical flow, whether stream-line or turbulent, there exists another entirely different type which may be called capillary or surface flow. In this latter type the liquid does not flow with its thermal condition unchanged. It is supposed, in fact, that close to a surface, only liquid without, or at most with only part of, its anomalous heat can pass. In sufficiently narrow capillaries, therefore, only flow of this kind can take place, and thus if two vessels A and B are connected by a narrow capillary and liquid is passed from A to B, heat will be evolved at A and absorbed at B. A direct experimental verification of this effect has been made by Daunt and Mendelssohn [29], who actually measured the fall in temperature when He II flows out of a tube through a small opening packed with powder. The action of the capillary might be compared with that of a semi-permeable membrane separating a solution from its pure solvent. The membrane allows only the solvent to pass freely and holds back the solute molecules; the capillary in a similar way appears to obstruct the passage of the excited atoms of He II. This analogy was first pointed out by Tisza [25].

H. London [5] has obtained an interesting equation based on the above ideas which connects the heat absorbed with the capillary flow. London assumes that in sufficiently narrow capillaries the heat conduction can be treated separately from the phenomenon which leads to the fountain effect. For instance, consider two vessels A and B connected by a narrow capillary. Let the temperatures and pressures in A and B be T, p and T_0, p_0 respectively. When $T>T_0$ we know that $p>p_0$. If, as London assumes, heat conduction can be treated separately, we can neglect its effect, for example by making the capillary narrow enough, and the liquid will then be in equilibrium under the above conditions. This assumption goes far beyond experimental facts as at present known, where reaction pressures have only been observed with an accompanying heat flow.

London considers the following reversible cycle. A volume of liquid passes from B to A by capillary flow and therefore does work $v(p-p_0)$. We neglect the effect of compressibility, which is not essential for the present purpose, in the interests of simplicity. The capillary flow will affect the temperature of the liquid in the two vessels and, to restore them to their original values, heat Q must be supplied to A and Q_0 taken from B. We now imagine a quantity of liquid v to be taken from A, cooled to temperature T_0, and returned to B. If the liquid is incompressible no work is done in this process. This completes the cycle, and the initial and final states are identical. We now express the fact that there is no total change in energy or entropy over the complete cycle. First, the energy condition is

$$Q=v(p-p_0)+Q_0+\rho v\int_{T_0}^{T} C_v\,dT. \qquad \ldots.(6)$$

The left-hand side represents the energy which has to be supplied from outside and the right-hand side is the heat which is given up and the work done by the liquid. The second condition (that the entropy change is zero) is

$$\frac{Q}{T}=\frac{Q_0}{T_0}+\rho v\int_{T_0}^{T}\frac{C_v\,dT}{T}, \qquad \ldots.(7)$$

where again the left-hand side is the entropy taken in and the right-hand side the entropy given out during the cycle.

We divide (6) by T_0 and subtract (7),

$$Q\left(\frac{1}{T_0}-\frac{1}{T}\right)=\frac{v}{T_0}(p-p_0)+\rho v\left\{\frac{1}{T_0}\int_{T_0}^{T}C_v\,dT-\int_{T_0}^{T}\frac{C_v\,dT}{T}\right\}. \qquad \ldots.(8)$$

In all practical cases T and T_0 lie between 1° and 2° and $T-T_0$ is of order at most, of 10^{-2} degree. Thus we may write $T=T_0+\Delta T$ and for uniformity of notation $p=p_0+\Delta p$. Retaining only first powers of ΔT, (8) becomes

$$\frac{Q}{v}=T\frac{\Delta p}{\Delta T}, \qquad \ldots.(9)$$

which is London's equation connecting the reaction pressure per degree with the heat absorbed per unit volume.

It is interesting to consider a special case, viz. when Q/T has its maximum possible value. Q/T is the entropy required to bring the liquid of volume v which flows through the capillary into A, up to the equilibrium condition at temperature T. If, now, we assume that the liquid flowing in the capillary is without its anomalous heat, we have as an upper limit to Q/T the value of the anomalous part of the entropy of v c.c. of liquid at temperature T. If we take the total entropy, the upper limit to Q/T will be rather over-estimated, so that we have, according to (9),

$$\left(\frac{\Delta p}{\Delta T}\right)_{\max}<J\rho S,$$

where S is the entropy in cal. per gm., $J=4{\cdot}18.10^7$ ergs per cal. is the mechanical equivalent of heat, and $(\Delta p/\Delta T)_{\max}$ is an upper limit to the reaction

pressure per degree according to this theory. At 1·5° K. $S=0{\cdot}06$ cal. per degree gm. Therefore $JS\rho=3{\cdot}7.10^5$ ergs/cm^3 At this temperature, according to the measurements of Allen and Reekie, the value of $\Delta p/\Delta T$ for a temperature difference $\Delta T=1{\cdot}25.10^{-3}$ is $3{\cdot}1.10^5$ dynes per cm^2

REFERENCES

(1) Allen and Misener. *Proc. Roy. Soc.* A, **172**, 467 (1939); *Nature, Lond.*, **141**, 75 (1938).
(2) Rollin. *Physica*, **3**, 269 (1936).
(3) Keesom and Keesom. *Physica*, **2**, 557 (1935); **3**, 359 (1936).
(4) Allen and Jones. *Nature, Lond.*, **141**, 243 (1938).
(5) London, H. *Proc. Roy. Soc.* A, **171**, 484 (1939); *Nature, Lond.*, **142**, 612 (1938).
(6) Daunt and Mendelssohn. *Proc. Roy. Soc.* A, **170**, 423 (1939).
(7) Keesom and Keesom. *Physica*, **2**, 128 (1934); **2**, 557 (1935).
(8) Schmidt and Keesom. *Physica*, **4**, 963 (1937).
(9) Keesom and Taconis. *Physica*, **5**, 161, 270 (1938).
(10) London, F. *Proc. Roy. Soc.* A, **153**, 576 (1936).
(11) Fröhlich. *Physica*, **4**, 639 (1937).
(12) Einstein. *S. B. preuss. Akad. Wiss.* 3 (1925); cf. also Fowler and Jones, *Proc. Camb. Phil. Soc.* **34**, 573 (1938), and London, F., *Phys. Rev.* **54**, 947 (1938) and *J. Phys. Chem.* **43**, 49 (1938).
(13) London, F. *Nature, Lond.*, **141**, 643 (1938).
(14) Allen and Reekie. *Proc. Camb. Phil. Soc.* **35**, 114 (1938).
(15) Kikoin and Lasarew. *Nature, Lond.*, **141**, 912 (1938).
(16) Burton. *Nature, Lond.*, **135**, 265 (1935).
(17) Wilhelm, Misener and Clark. *Proc. Roy. Soc.* A, **151**, 342 (1935).
(18) Keesom and Macwood. *Physica*, **5**, 737 (1938).
(19) Burton. *Nature, Lond.*, **142**, 72 (1938).
(20) Kapitza. *Nature, Lond.*, **141**, 75 (1938).
(21) Cf. Prandtl and Tietiens. *Applied Hydro- and Aeromechanics.*
(22) Keesom, Keesom and Saris. *Physica*, **5**, 281 (1938).
(23) Allen, Peierls and Uddin. *Nature, Lond.*, **140**, 62 (1937).
(24) Kürti and Simon. *Nature, Lond.*, **141**, 207 (1938).
(25) Tisza. *Nature, Lond.*, **141**, 913 (1938); also *C.R. Acad. Sci., Paris*, **207**, 1035, 1186 (1938).
(26) Giauque, Stout and Barieau. *J. Amer. Chem. Soc.* **61**, 654 (1939).
(27) Jones. *Proc. Camb. Phil. Soc.* **34**, 253 (1938).
(28) Michels, Bijl and de Boer. *Physica*, **5** (1938).
(29) Daunt and Mendelssohn. *Nature, Lond.*, **143**, 719 (1939).

HEAT

By R. W. POWELL, Ph.D.
Physics Department, The National Physical Laboratory,
Teddington, Middlesex

§ 1. INTRODUCTION

Various sub-sections of the subject of heat have been dealt with in earlier *Reports*, and the main purpose of the present article is to describe the recent advances in some of these sections. There has, however, been a request for the inclusion of an article on "Measurement of High Temperature in Industry", and, despite the author's limited industrial experience, an attempt has been made to comply with this request.

§ 2. THE MEASUREMENT OF HIGH TEMPERATURE IN INDUSTRY *

A paper [1] entitled "Pyrometric practice" was published by Foote, Fairchild and Harrison of the staff of the Bureau of Standards in 1921. This paper gave a very comprehensive account of the subject of high-temperature measurement, and dealt with the technical features of each type of instrument, with a view to providing information which would ensure a suitable instrument being selected for a given type of measurement, and would enable it to be used most efficiently. Today, the paper is still a standard reference work in which an answer can be found to most queries regarding high-temperature measurement by means of resistance thermometers, thermocouples and pyrometers of either the optical or total radiation type. It should therefore suffice for the present article to deal only with subsequent developments in the subject.

Probably the most important development has been the adoption by general agreement of the international temperature scale [2], extracts from which were given in Volume I of these *Reports*. As well as defining the various fixed points of the scale, the means by which the scale is to be realized at high temperatures were also specified and are repeated here.

From the ice point to 660° c., the temperature, t, to be deduced from the resistance R_t of a standard platinum resistance thermometer by means of the formula, $R_t = R_0(1+At+Bt^2)$, the three constants to be evaluated from resistances measured at the ice, steam and sulphur points. From 660° to 1063° c. the temperature to be deduced from the electromotive force e of a standard platinum *vs.* 90 per cent platinum, 10 per cent rhodium thermocouple, according to the formula $e=a+bt+ct^2$, where the constants are determined from observations at the freezing points of antimony (as measured by a resistance thermometer), silver (960·5° c.) and gold (1063° c.). Above the gold point the temperature to be determined by means of the ratio of the

* The author is indebted to Dr. F. H. Schofield for assistance in the preparation of this section.

intensity J_2 of monochromatic visible radiation of wave-length λ cm., emitted by a black body at the temperature t, to the intensity J_1 of radiation of the same wave-length emitted by a black body at the gold point, by means of the formula

$$\log_e \frac{J_2}{J_1} = \frac{C_2}{\lambda}\left[\frac{1}{1336} - \frac{1}{(t+273)}\right].$$

The constant C_2 is taken as 1·432 cm.-degrees and the equation is valid if $\lambda(t+273)$ is less than 0·3 cm.-degrees.

The standardizing laboratories have to develop and maintain equipment for the interpolation of the scale with the highest precision. Hall [3] has described the precautions to be taken in the construction and use of precision resistance thermometers, and Barber [4] has published details of a sulphur-point bath with internal electric heating, which has proved an efficient piece of apparatus for their calibration at the upper fixed point. Quite recently, de Selincourt [5] has completed an investigation which concerned the degree of reproducibility obtainable with the standard platinum thermocouple at the three fixed points within its recommended range. In a few instances an accuracy of $\pm 0{\cdot}01°$ C. was obtainable but $\pm 0{\cdot}1°$ C. was more usual, the readings being found to be susceptible to small amounts of strain within the thermocouple wires. The latter value represents the order of agreement reached in an intercomparison [6] between the standardizing laboratories of Germany, the United States and Great Britain over the same temperature range.

At higher temperatures it has been considered necessary to determine the melting points of palladium and rhodium, that they might serve as secondary fixed points. Such a secondary point can be used for evaluating temperatures in its neighbourhood instead of the more lengthy procedure of always working back to the gold point in accordance with the strict definition of the international temperature scale. The freezing point of palladium, 1555° C., and the melting point of tungsten, 3400° C., were the only such points originally specified above the gold point. The close agreement obtained for palladium itself and for the secondary metals so far examined at the three standardizing laboratories just referred to can be seen from the data given in table 1.

Table 1. *Freezing points* (°C.) *of metals suitable for use as secondary standards above the gold point*

Metal	Laboratory at which determined		
	N.B.S.	N.P.L.	P.T.R.
Palladium	$1553{\cdot}6 \pm 0{\cdot}5$ [7]	$1554{\cdot}4 \pm 1{\cdot}0$ [8]	1556 [9]
Platinum	$1773{\cdot}5 \pm 1$ [10]	$1773{\cdot}3 \pm 1$ [11]	$1773{\cdot}8 \pm 1$ [12]
Rhodium	1966 ± 3 [13]	1966 ± 3 [14]	—

For these determinations, very pure samples of the metals have been melted by induction heating, and the temperatures of black-body radiators immersed in the ingots have been measured by optical pyrometers used in conjunction with a rotating sector which reduces the brightness at the higher temperature to that of the gold point. In the latest work with rhodium [14] the required reduction amounted to about 700 to 1, and it has been necessary to perform this reduction in two stages, first to 1270° C. and then to 1063° C.

Optical pyrometers suitable for such precision measurements have also been described by Forsythe [15] and Cunnold [16] [17], whilst Stott [18] has shown that the range of a pyrometer of the disappearing filament type can be extended by some 80° C. at the lower end if a slit set parallel to the filament is used instead of the usual circular aperture.

Having now briefly outlined the temperature scale and the means whereby it is maintained we can turn our attention to the wider sphere of the measurement of high temperature in industry. Here we find that the principal instruments are the same as those already considered, namely, resistance thermometers, thermocouples, optical and total radiation pyrometers, but that their use is not invariably restricted to the ranges of temperature specified in the international temperature scale. Resistance thermometers are seldom used at temperatures above about 500° C., but thermocouples of various kinds can be employed for measurements at all temperatures likely to occur in industrial processes.

Owing to their greater cost platinum thermocouples find their most useful application in the region of temperature above that at which base-metal thermocouples are available, say from 1000° C. to 1100° C. upwards. The platinum thermocouple is available in two forms consisting of pure platinum against an alloy of platinum with either 10 or 13 per cent of rhodium. There has been of recent years a notable tendency to increase the upper limit for platinum thermocouples, as is shown by the fact that the calibration data are now commonly given to 1550° C. in place of 1300° C. as formerly. The several makers reproduce their calibration data from year to year to about $\pm 1°$ C. at 1000° C. and $\pm 3°$ C. at 1500° C. The National Bureau of Standards has published e.m.f.-temperature tables [19] for both 10 and 13 per cent thermocouples from 0° C. to about 1700° C., and it is considered that the use of these tables, together with deviation curves obtained by calibrations at two points of about 660° C. and 1200° C., gives a calibration which is not in error by more than 3° C. to 1500° C. Similar standard tables have been published for many of the base-metal thermocouples available for use at lower temperatures. These include iron *vs.* constantan, from −200° to 1000° C.[20], copper *vs.* constantan, from −200° to 400° C.[20], and chromel *vs.* alumel, from 0° to 1400° C.[21]. The upper temperature marks the limit to which each thermocouple can be used. The papers describe the calibration procedure, and contain information regarding the technique to be adopted, and the accuracy likely to be obtained with the various thermocouples. A survey giving the e.m.f.s of all metals which can be considered for use as thermoelements has recently been published by Schulze [22].

The optical pyrometer can be used for the measurement of temperatures from about 750° C. upwards. This method suffers from the complication that, unless the conditions are those of a black body, a knowledge of the emissivity of the surface on which the pyrometer is sighted is required before the true temperature can be evaluated.

Photoelectric cells have been incorporated in the more recent patterns of radiation pyrometers, and instruments of this type are described by Zetzmann [23]. One such instrument has been introduced [24] which operates in the same way as a camera. This so-called "Snap" pyrometer is readily sighted on any portion of the furnace by means of a reflection viewer, and, during exposure, the shutter remains open for a second or so and allows the radiation to fall on a photocell. The resultant current is integrated by the indicator, which moves over a scale calibrated from 1000° C. upwards.

Papers describing the foregoing normal methods of high-temperature measurement have been published by Sosman [25], Marsh and Dike [26], Letort [27], Moser and Tingwaldt [28], and Ribaud [28a] has devoted a book to the subject. Sosman [29] has also dealt with the use of pyrometric cones for indicating the firing temperatures of refractories, and, in the same paper, has described some methods of high-temperature control which have been introduced for the preservation of refractories in open hearth furnaces. These methods can well be mentioned here as illustrating instances where the ingenuity of the physicist has enabled manual control to be replaced by automatic devices.

Owing to its reaction with iron oxide the refractory roof of an acid open-hearth furnace is liable to melt if heated above 1650° C., and, since the charge of steel has to be taken to about 1585° C. for pouring purposes, it follows that there is only a safety margin of 65° C. The innovation has been to observe the internal roof temperature through a hole in the brick wall of the furnace by means of either a total-radiation pyrometer or a partial-radiation instrument. In the former instrument the energy from the hot surface is received by a thermocouple, and in the latter by a vacuum photoelectric cell which absorbs part of the infra-red radiation and gives rise to a measurable current. Water cooling is used to screen the sensitive elements from other sources of radiation, and in each instrument the temperature record is obtained by means of a recording potentiometer. If the calibration is done directly in the furnace under normal flame conditions, the instrument can be set for automatic control with the normal fuel. Where a furnace of the regenerative type is used, a further refinement has been to install another radiation pyrometer which automatically reverses the direction of gas flow through the furnace. In such a furnace, the products of combustion are led over a loosely packed checker-work of refractory brick inside the regenerator chamber. These bricks become heated and after the reversal they serve to preheat the air supplied for combustion. It is important to guard against overheating the bricks as this would cause softening, and would ultimately lead to collapse. Thus automatic reversal by means of a pyrometer sighted directly on the bricks

in each regenerator chamber ensures smooth working with freedom from this possible source of trouble.

A control device of interest to fuel engineers is the " pyromaster " which has been described by Dupen [30] and applied to the outlet temperature of a producer gas plant. In this instance the gas temperature is measured by a chromel alumel thermocouple connected to a recording potentiometer, subsidiary apparatus being so arranged that departures from the balance position operates relays regulating the rate of fuel supply to the plant.

A problem of actual temperature measurement which has long awaited a satisfactory solution has been that of the temperature of the liquid steel itself when in the furnace, and at various stages of the pouring operation. Until recently, the measurement was usually attempted by means of an optical pyrometer, but as the surface emissivities were not known, and varied in different regions, the temperatures could only be obtained approximately since, at the high temperatures involved, the correction (using a wave-length of $0{\cdot}65\mu$) varies from about 35° C. for an emissivity of 0·8 to 300° C. for an emissivity of 0·2. Various estimates have given the correction when sighting on the stream of metal issuing from the furnace as from 60° to 190° C. Blaurock [31] stated that the temperature of molten steel could be obtained to within 15° C. if measurements were made using first a red and secondly a green filter in the pyrometer, the true temperature then being derived from a curve representing the true temperature minus the apparent temperature with the red filter, plotted against the difference between the apparent temperatures obtained with the red and green filters. Naeser [32] has described a special type of pyrometer which includes a colour pyrometer, a comparison lamp, and an adjustable tinted neutral wedge, with which he claims that steel temperatures can be determined to $\pm 10°$ C.

In this country attention has been given to the problem bya Sub-Committee of the Iron and Steel Institute, and in their latest report [33] they have recommended the adoption of a " quick immersion " thermocouple for the general measurement of liquid steel temperatures. This recommendation is based on the work of Schofield and Grace [34], who felt that a method depending on direct contact with the liquid steel in the furnace was obviously to be desired. Such readings could be obtained by means of a platinum thermocouple, heavily sheathed in the usual manner, but the wastage of sheath and platinum rendered its routine application impracticable. Schofield and Grace found, however, by small-scale laboratory experiments, that a platinum thermocouple, lightly sheathed in silica, could be plunged into liquid steel so as to give a reading of temperature in a few seconds and be withdrawn intact. They have now described in detail means of applying this " quick immersion " technique for the routine measurement of steel temperatures in open-hearth and electric furnaces, whether of the acid or basic types, and also for such measurements at other stages in the passage of the steel from the furnace to the mould, i.e. the launder, ladle, trough, and the mould itself. Apparently the thermocouple readings are unaffected by electrical interference

from induction furnaces under load, operating at frequencies of 500, 1500 and even more cycles per second.

It is interesting to note that the highest temperature of steel (1700° C.) recorded by a platinum thermocouple worked on this plan is not far removed from the melting point of platinum. The silica sheath requires replacement after each immersion, but it is considered that the thermocouple wire (0·5 mm. diameter platinum *vs.* platinum 13 per cent rhodium) can be used for ten immersions, and probably a considerably greater number can be tolerated without risk of serious contamination.

In addition to designing the thermocouple equipment to give equilibrium readings as quickly as possible it is important to use an indicating system which responds rapidly. Details of four indicators are given by Schofield and Grace: (1) a pivoted millivoltmeter; (2) a quick-period reflecting millivoltmeter; (3) an interpolation potentiometer; and (4) a slide-wire potentiometer. The authors favour the last-mentioned method, with its freedom from errors due to change in galvanometer sensitivity and circuit resistance, although the time taken to get a reading may be a second or so longer than with (2) or (3).

Keinath [35] has dealt with the subject of temperature measurement and control in steel-hardening furnaces. Here base-metal couples are used and it is suggested that the development of large temperature differences between thick and thin portions of the specimen can be prevented, particularly in the neighbourhood of the steel transformation zone, if two thermocouples are used, one in the specimen and the other in the furnace, and these are arranged to control the rate of heating so that there is a constant temperature difference between the two thermocouples.

In the ceramic industry, temperatures are still determined by means of pyrometric cones. These were introduced by Seger, who utilized the fact that the fluidity of a fused silicate increases rapidly with temperature, and is very dependent on composition, to make a series of cones which "squatted" at fairly definite temperatures. Such cones have since been improved and standardized by Orton. They are not true pyrometers, since the squatting depends on time as well as temperature, but this is not a serious defect under the conditions in which they are used, and the cones have the merit that they are small and can be placed exactly where the temperature measurement is required. About 60 cones composed of various minerals such as kaolin, quartz, felspar, talc, calcium carbonate, ferric oxide, and so on, cover the range 600° to 2015° C.[35a] Rea [36] has recently suggested that a series having definite melting temperatures over a range from 500° to 1000° C. can be made from eutectoid combinations of various salts.

Although pyrometric cones are very simple to use, the determination of their true squatting temperatures presents many difficulties, particularly in the case of the more refractory cones, since, in addition to heating the cones at a controlled rate of about 3° C. per minute, it is essential that the atmosphere of the furnace should be oxidizing. Stott [37] has described such a determination of the squatting temperatures of cones nos. 26 to 35 (1564° to 1755° C.).

The variations between cones of the same batch were found to be small, and it was considered unlikely that the squatting points of the cones were in error by much more than 5° C. with respect to the international temperature scale. It thus appears that the precision of a given batch of these cones used under specified conditions is little less than that of a good optical pyrometer.

The subject of flame-temperature measurement has been adequately dealt with in Lewis and von Elbe's recent book [38] *Combustion, Flames and Explosions of Gases*, and an appendix has been included which lists numerous flame temperatures. Most of the measurements have been made by means of the spectral-line reversal method. It might however be noted that Hübner [39] has since described a modification by which it becomes possible to measure temperatures which are considerably higher than that of the reversing radiation, brightness measurements of the line-source being extrapolated to give the temperature of reversal.

From time to time new methods are suggested for measuring high temperatures, and, though all may not find a definite use in industrial processes, it seems appropriate to refer to a few of these before bringing this section to a close, as they may suggest new avenues for exploration when attempting to solve some particular problem of temperature measurement.

Grüss [40] has suggested the use of resistance thermometers which depend on the large decrease in electrical resistivity which occurs in refractory materials at temperatures above about 800° C. A sintered corundum thermometer of this type is described, having platinum electrodes fused on by means of porcelain, which has been found to give consistent results over the range 800° to 1100° C. As there is definite departure from Ohm's law, the thermometer would need to be calibrated and used under similar conditions of current flow.

For the measurement of temperatures in the visible region Hase [41] some years ago suggested a photographic method. A knowledge of the emissivity of the surface is required in order to interpret the records, but it is interesting to note that Neubert [42], by infra-red photography, has been able subsequently to measure temperatures from 250° C. upwards in this way. The method may ultimately prove of use for measuring the temperature distribution of moving parts of machinery, or for other problems where it may not be convenient to attach thermocouples.

Another method which has recently become available for such a purpose, and one in which it is not necessary to be able to observe the surface in question, consists in the use of temperature-sensitive paints. Penzig [43] has described a number of such paints which undergo marked changes in colour on heating to various temperatures. The temperatures which can be detected in this way are 30, 60, 65, 110, 140, 145, 155, 175, 220, 230, 275, 290, 340 and 440° C. There is some dependence on time of heating, but the temperatures are considered to be defined to within 5° C. As the changes are irreversible, the use of a suitable paint enables a permanent record to be obtained, from which it is possible, within certain limits, to ascertain the surface temperature above or below which a metal part has been heated.

REFERENCES

(1) Foote, P. D., Fairchild, C. O. and Harrison, T. R. *Tech. Pap. Bur. Stand.* no. 170, Feb. 1921.
(2) *Trav. Bur. int. Poid. Mes.* **18**, 94 (1927); and **19** (1934), revised text.
(3) Hall, J. A. *J. Sci. Instrum.* **10**, 4 (1933).
(4) Barber, C. R. *J. Sci. Instrum.* **14**, 227 (1937).
(5) de Selincourt, M. *Proc. Phys. Soc.* **51**, 695 (1939).
(6) Roeser, F. W., Schofield, F. H. and Moser, H. A. *Ann. Phys., Lpz.*, **17**, 243 (1933); *Bur. Stand. J. Res., Wash.*, **11**, 1 (1933).
(7) Fairchild, C. O., Hoover, W. H. and Peters, M. F. *Bur. Stand. J. Res., Wash.*, **2**, 931 (1929).
(8) Schofield, F. H. *Proc. Roy. Soc.* A, **155**, 301 (1936).
(9) Hoffmann, F. and Meissner, W. *Ann. Phys., Lpz.*, **60**, 201 (1919).
(10) Roeser, W. F., Caldwell, F. R. and Wensel, H. T. *Bur. Stand. J. Res., Wash.*, **6**, 1119 (1931).
(11) Schofield, F. H. *Proc. Roy. Soc.* A, **146**, 792 (1934).
(12) Hoffmann, F. and Tingwaldt, C. *Phyz. Z.* **35**, 434 (1934).
(13) Roeser, W. F. and Wensel, H. T. *Bur. Stand. J. Res., Wash.*, **12**, 519 (1934).
(14) Barber, C. R. and Schofield, F. H. *Proc. Roy. Soc.* A, **173**, 117 (1939).
(15) Forsythe, W. E. *J. Opt. Soc. Amer. & Rev. Sci. Instrum.* **10**, 19 (1925).
(16) Milford, M., Bracey, R. J., Cunnold, F. A. and Egerton, A. C. *J. Sci. Instrum.* **12**, 80 (1935).
(17) Cunnold, F. A. *Proc. Roy. Soc.* A, **152**, 64 (1935).
(18) Stott, V. H. *J. Sci. Instrum.* **14**, 370 (1937).
(19) Roeser, W. F. and Wensel, H. T. *Bur. Stand. J. Res., Wash.*, **10**, 275 (1933).
(20) Roeser, W. F. and Dahl, A. I. *Bur. Stand. J. Res., Wash.*, **20**, 337 (1938).
(21) Roeser, W. F., Dahl, A. I. and Gowens, G. J. *Bur. Stand. J. Res., Wash.*, **14**, 239 (1935).
(22) Schulze, A. *Metallwirtsch., Metallwiss.*, **18**, 249, 271 and 315 (1939).
(23) Zetzmann, J. *Arch. tech. Messen.* **4**, T. 116 (1934).
(24) Camb. Inst. Co., Ltd. *J. Sci. Instrum.* **15**, 277 (1938).
(25) Sosman, R. B. *Industr. Engng Chem.* **23**, 1369 (1931).
(26) Marsh, K. and Dike, P. H. *Metal Progress*, **30**, 167 (1936).
(27) Letort, Y. *Céramique*, **40**, 167 (1937).
(28) Moser, H. and Tingwaldt, C. *Die Physik*, **6**, 1 (1938).
(28 *a*) Ribaud, G. *Mesure des Températures.* (Armand Colin, Paris 1936.)
(29) Sosman, R. B. *J. Amer. Ceram. Soc.* **21**, 37 (1938).
(30) Dupen, C. *J. Inst. Fuel*, **12**, 40 (1938).
(31) Blaurock, F. *Arch. Eisenhüttenw.* **8**, 517 (1934–5).
(32) Naeser, G. *Mitt. K.-Wilh.-Inst. Eisenforsch*, **18**, 21 (1936).
(33) *Eighth Report on the Heterogeneity of Steel Ingots.* Special Report no. 25, Section vii, p. 235 (1939).
(34) Schofield, F. H. and Grace, A. *Eighth Report on the Heterogeneity of Steel Ingots.* Special Report no. 25, Section vii, p. 239 (1939).
(35) Keinath, G. *Arch. tech. Messen.* no. 84, T. 71 (1938).
(35 *a*) Fairchild, C. O. and Peters, M. F. *J. Amer. Ceram. Soc.* **9**, 701 (1926).
(36) Rea, R. F. *J. Amer. Ceram. Soc.* **21**, 98 (1938).
(37) Stott, V. H. *Trans. Ceram. Soc.* **38**, 341 (1939).
(38) Lewis, B. and von Elbe, G. *Combustion, Flames and Explosions of Gases.* (Camb. Univ. Press, 1938.)
(39) Hübner, H. J. *Ann. Phys., Lpz.*, **33**, 52 (1938).
(40) Grüss, H. *Arch. tech. Messen.* no. 78, T. 161 (1937).
(41) Hase, R. *Z. tech. Phys.* **13**, 410 (1932).
(42) Neubert, P. *Arch. Wärmew.* **19**, 29 (1938).
(43) Penzig, F. *Z. Ver. dtsch. Ing.* **83**, 69 (1939).

§ 3 TEMPERATURE MEASUREMENTS AT NORMAL AND LOW TEMPERATURES

A recent development of applied physics which has required special methods for the measurement and recording of temperature has been that in which the properties of the upper atmosphere are explored by sounding balloons. These balloons carry small wireless transmitters which send out signals arranged to give intermittent records of the atmospheric pressure, temperature and humidity. Thomas [44] has given some account of the various principles used in such radio-meteorography and has dealt particularly with apparatus used at the National Physical Laboratory. For temperature measurement the normal type of bi-metallic strip was found to possess too much mechanical hysteresis. The method adopted for this measurement depends on the change in length of a 38 S.W.G. brass wire held in an invar frame, and kept under slight tension by an invar spring to which is attached a "Stalloy" iron plate. As this plate moves, in consequence of the change of length of the brass wire, it varies the gap of an iron-cored coil and alters its inductance. The initial sag of the wire is adjusted at room temperature so that the wire becomes taut at about −80° C. The wire is then soldered at each end, care being taken that no overheating of the wire occurs. Over the temperature range 20° to −80° C. the movement of the iron plate is about 0·1 inch, and this is accompanied by a frequency change of some 200 cycles per second. Since a change in frequency of 0·5 cycles per sec. can be observed, the temperature can be measured by this apparatus to about 0·1° C. at 20° C., and about 0·6° C. at −80° C.

The method of temperature measurement employed in radio-meteorography at the National Bureau of Standards has been described by Diamond, Hinman and Dunmore [45] and also by Craig [46]. Here a capillary-tube thermometer has been developed which, besides having the advantage of freedom from mechanical parts, responds more quickly to changes in temperature than do bi-metallic thermometers of the usual type. The instrument consists of a glass tube about 8 cm. long, 0·75 mm. bore, and 0·4 mm. wall thickness bent into the form of a U and having a small glass bulb formed at each end. The tube is filled with an electrolyte and copper electrodes are sealed into each end by means of small rubber bungs. Several experiments have been carried out to determine a suitable electrolyte. The one finally used consists of a mixture of hydrochloric acid, cuprous chloride and alcohol. With this mixture ample cuprous chloride is taken up by the acid and alcohol to eliminate polarization almost completely, and this results in the resistance being nearly independent of the current used. A wide range of resistances can be obtained by suitable choice of the proportions of the three ingredients. A mixture of 24 per cent by volume of hydrochloric acid and 76 per cent of ethyl alcohol to which has been added 2·7 gm. of cuprous chloride gave resistances of 30,000 ohms at 30° C. and 56,000 ohms at −50° C. and remained fluid to −78° C. Temperatures can be accurately measured in this way to within 1° C. even at the lowest temperature.

Craig points out that the use of such capillary-tube thermometers is not

limited to the field of radio-meteorography. Their high temperature coefficient of resistance and stability suggests that they should be suitable for temperature control. Some preliminary experiments in this direction are described, the temperature of a water-bath being controlled to $\pm 0{\cdot}005°$ C. by such a thermometer used with alternating current in conjunction with a radio amplifier and small thyratron. Details of another thermoregulator which also employs a resistance thermometer and a phase-shifting thyratron circuit and which maintains a water bath to $\pm 0{\cdot}003°$ C. have been given by Sturtevant [47]. A very simple method of thermostat control has been described by Josten [48] in which the temperature variant is a liquid in a vertical tube. This tube is transparent to light except in the region of the meniscus, and thus, as the meniscus passes across the beam, it can easily be arranged for the reduction in photoelectric current to operate a relay controlling the power supply to the enclosure.

The accurate measurement of the temperature of heated surfaces exposed to the air presents considerable difficulties in many instances. It is essential not to alter the surface conditions during measurement. Two methods which involve measurements of the energy radiated from the surface have been suggested during the past year. Roose [50] suggests that the wall temperature of a room can be derived from observations taken with two spheres, suspended about a yard above the floor and maintained electrically at 36·5° C. One sphere is nickel-plated and has an emissivity of about 0·04 and the other is painted black and has an emissivity of 0·965. The wall temperature is calculated in terms of the difference between the two energy measurements. The second radiometric device has been shown by Hardy and Soderstrom [51] to be capable of measuring black-body temperature to $\pm 0{\cdot}01°$ C., and has been used for measuring the skin temperatures of human beings. Independent tests have shown that the skin has an emissivity within 2 per cent of that of a black body. The instrument simply consists of a thermopile calibrated against a black body of known temperature.

For other reasons, the measurement of the temperature of a very rapidly flowing stream of air is no simple matter, for in this case a certain amount of heat is likely to be produced by the air sweeping over the thermometer Some experiments in this connection have been carried out by Hilton [52]. He points out that a circular cylinder placed at right angles to the direction of the air flow is a most unsatisfactory shape for a thermometer, since its temperature varies from nearly the adiabatic-compression temperature at the front to nearly the free-stream temperature at the rear. In fact a thermometer of the normal shape would give a more satisfactory reading if placed end-on with reference to the air stream. Failing a thermometric device which can be moved in the direction of the air stream and at the same speed, the best type is considered to be a flat plate, with a thermocouple attached in the region where the boundary-layer flow is laminar. In this region, Hilton's experiments proved that the increase in temperature could be calculated by means of an equation originally suggested by Polhausen.* The subject has also been treated theoretically by Kroujiline [53].

* See the article by Fage, p. 270. [Editor.]

For the calibration of thermometers in the neighbourhood of atmospheric temperatures, use is often made of sodium sulphate decahydrate which undergoes a transformation at 32·384° C. Washburn and Clem [54] have now made an accurate determination of the transformation temperature of sodium sulphate heptahydrate. This occurs at 23·465±0·004° C., and as the salt can be easily purified, its transition temperature is also recommended for use as a secondary fixed point.

A new form of gas thermometer has been proposed by Woodcock [55] for use in the difficult region between the temperatures of liquid helium and liquid hydrogen. The thermometer is unusual in that the volume of gas which remains at atmospheric temperature is much greater than that in the thermometer bulb proper, which is immersed in the low-temperature bath. If these two volumes are in the ratio of 10 to 1 the sensitivity at temperatures of 20°, 15°, 10° and 5° K. is respectively 4·4, 5·4, 6·8 and 8·9 times that of a gas thermometer of the usual pattern in which the external volume is kept relatively small. This sensitivity permits temperatures below 20° K. to be read to within 0·1° by means of an ordinary mercury manometer, and yet the pressure in the instrument is only of the order of 1 atmosphere at room temperature.

REFERENCES

(44) Thomas, H. A. *Proc. Roy. Soc.* A, **167**, 227 (1938).
(45) Diamond, H., Hinman, Jr., W. S. and Dunmore, F. W. *Bur. Stand. J. Res., Wash.*, **20**, 369 (1938).
(46) Craig, D. N. *Bur. Stand. J. Res., Wash.*, **21**, 225 (1938).
(47) Sturtevant, J. M. *Rev. Sci. Instrum.* **9**, 276 (1938).
(48) Josten, G. W. *Industr. Engng. Chem. Analy. Edn.* **10**, 165 (1938).
(49) Shepherd, M. *Bur. Stand. J. Res., Wash.*, **21**, 831 (1938).
(50) Roose, H. *Gesundheitsing.* **61**, 325 (1938).
(51) Hardy, J. D. and Soderstrom, G. F. *Rev. Sci. Instrum.* **8**, 419 (1937).
(52) Hilton, W. F. *Proc. Roy. Soc.* A, **168**, 43 (1938).
(53) Kroujiline, G. *Techn. Phys. U.S.S.R.*, **4**, 73 (1937).
(54) Washburn, E. R. and Clem, W J. *J. Amer. Chem. Soc.* **60**, 754 (1938).
(55) Woodcock, A. H. *Canad. J. Res.* **16**, 133 (1938).

§ 4. HYGROMETRY

As mentioned in the previous section, measurements of the relative humidity of the upper atmosphere as well as of its pressure and temperature are made in radio-meteorography. Hair hygrometers were at first used [45] to record the changes in humidity, and owing to the serious time-lag in the response of instruments of this type, the accuracy of the relative-humidity measurement was considerably less than that of temperature. As this lag increases at lower temperatures, the accuracy of the hair hygrometer becomes progressively less as the altitude is increased. In practice, a small metal cam is swung about its shaft by two hair elements operating in series, and a pivoted wire-wound resistance is held in contact with the cam by a spring. Thus, as the humidity alters from 0 to 100 per cent, the contact of the cam on the resistance is moved from one end to the other.

Dunmore [56] has recently described a new type of electric hygrometer which appears to possess several advantages over the hair hygrometer for

this type of work. The new unit is in the form of a resistor which is connected to the input of the audio-oscillator used in the radio-meteorograph. The variations of resistance produced by changes of humidity are translated into audio-frequencies which are recorded at the ground station on a graphic frequency recorder. The sensitive resistor consists of a thin-walled glass tube, having its outer glass surface roughened and wound with a dual winding of 38-gauge tinned copper wire. The variable resistance is the glass surface, the wires serving as electrodes. Kimble flint glass in the form of a tube 10 mm. in diameter, 12·7 cm. long and 0·3 mm. wall thickness has been found to make a satisfactory unit when wound with 20 turns of wire per inch. The units are coated with solutions of lithium chloride, as this procedure enables the change in resistance with humidity relative to the change in resistance due to temperature to be varied. Thus an uncoated unit shows the least dependence on temperature, whilst an instrument freshly coated with 3 per cent lithium chloride solution would be used in the stratosphere, where the temperature changes are smaller, and the quickest possible response to humidity changes is required. In a comparison of an instrument of this new type with one of the hair pattern, the former responded to sudden changes in humidity about two minutes sooner. The time-lag constant of the dual-coil hygrometer is of the same order as that of the capillary-tube thermometer and it is evidently much superior to the hair hygrometer and allows more exact records of the humidity of the upper atmosphere to be obtained. Calibration has to be carried out frequently, as the instrument suffers from both long- and short-period ageing changes. A footnote to the paper, however, indicates that the long-period ageing is due to a resistance film being built up at the wire-glass contact, and that this can be eliminated by using platinum wire instead of tinned copper.

The accurate measurement of relative humidity at low temperatures, even under laboratory conditions, is by no means an easy matter, and this new device will no doubt become more widely used. Ebert [57] has published a review of recent contributions to the subject of hygrometry and has dealt separately with humidity measurements at temperatures below 0° C.[58]. In the latter paper, preference was given for the dew-point and wet-and-dry-bulb methods of measurement, and the need was stressed for further experimental research in this temperature region where a comparison of different methods had shown lack of agreement between the results obtained.

The dew-point method is an absolute method of relative humidity measurement which has a sound physical basis. From an observation of the temperature at which condensation occurs, the corresponding saturation vapour pressure is derived. The ratio of this pressure to the saturation vapour pressure at the temperature of the air, or enclosure, gives the required relative humidity of the air. Care should always be taken to avoid too heavy a deposit of dew so that the temperatures of appearance and disappearance of dew lie close together. When working in enclosures at low temperatures this point should be given still greater care, and a very sensitive method used to detect the presence of the dew, for the actual water-vapour content of the

air is then lower, and the deposition of a heavy film might cause a measurable lowering of the vapour content of the sample of air under test. In addition, since the temperature of the air-liquid interface cannot be observed, it is also essential that the temperature drop across the dew film shall be as small as possible, i.e. the film of dew should be as thin as possible, and the temperature of the metal on which the dew is formed should be measured by means of a thermojunction located very close to the surface.

An apparatus for the measurement of the dew point at low temperatures was described a few years ago by Griffiths and Awbery [59]. It consisted essentially of a polished metal disc, cooled internally by the circulation of pre-cooled paraffin, a copper constantan thermocouple being used to measure the temperature of the disc. A comparison of the values obtained gravimetrically and by means of this instrument over the range $-5{\cdot}8°$ to $-16{\cdot}4°$ C. indicates the percentage relative humidity given by the dew-point method to be 3 per cent low on the average. In commenting on this result, the authors suggest that the values assumed for the saturation moisture content of air at these low temperatures may be in error.

It appears that the presence of the dew can be detected with greater sensitivity by means of photoelectric devices than by the eye. Some years ago Moss [60] described a dew-point instrument in which oblique illumination of the mirror was used, the mirror being automatically maintained at the dew point by such a photoelectric apparatus. Quite recently Hixson and White [61] have dealt with the accurate determination of dew-point temperatures. They describe an apparatus in which the dew is formed on a chromium-plated brass disc, cooled by a thermostatically controlled water-flow. The gas under test circulates slowly over the mirror surface, and light incident at an angle of about 10° to the mirror surface is reflected on to a photocell, which controls the grid of a three-element vacuum-tube amplifier. It was found best to work with slowly increasing mirror temperatures, the disappearance of dew being accompanied by a change in slope of the curve of temperature against plate current. No information is given regarding the performance of this instrument at low temperatures or humidities, but at normal temperatures it is claimed to have a precision of $\pm 0{\cdot}01°$ C. at 90 per cent relative humidity and $\pm 0{\cdot}05°$ C. for saturated air. Winkler [62] has since described a dew-point hygrometer for use at low temperatures which enabled humidities within the range 60 to 95 per cent to be measured to $\pm 0{\cdot}5$ per cent down to $-15°$ C. This degree of accuracy was obtained with visual observation when the mirror was slowly cooled at the rate of about 2° C. per hour by the circulation of a stream of non-freezing liquid. As a means of further increasing the accuracy of the instrument it is proposed to incorporate a photoelectric detector and to use still lower rates of cooling. A portable instrument which comprises a dew-point hygrometer and a potentiometer (for measuring the temperature of the thermocouple) has been described by Stack [63]. It is intended for making routine determinations of the moisture contents of furnace gases.

When the relative humidity is derived from measurements made with

a psychrometer of the wet-and-dry-bulb type, use has to be made of suitable tables, or of a formula, such as that originally proposed by Sprung :

$$e'-e=0{\cdot}5\,(\theta-\theta')P/755,$$

where θ and θ' are the dry- and wet-bulb temperatures, e the vapour pressure of water in the atmosphere of total pressure P, and e' the saturation vapour-pressure at temperature θ'. The tables published [64] for use with the Assmann psychrometer are based on this equation. It has, however, been a matter of discussion as to whether the numerical coefficient can invariably be taken as 0·5. In 1932 Whipple [65] analyzed and tabulated the theoretical and experimental data then in existence for the determination of this so-called psychrometric constant. Considerable variations were revealed, but these were shown to have comparatively little effect on relative humidity determinations carried out at air temperatures of 30° C. and upwards. Awbery and Griffiths [66] as a result of their later work considered that there was evidence above 0° C. for the " constant " being a linear function of the wet-bulb temperature ; below 0° C. it could be neither constant nor a function of wet-bulb temperature only. In view of this complication Awbery and Griffiths elected to smooth their observations at low temperatures graphically, rather than to attempt to fit them to an equation. A table is however included in the paper which indicates considerable variations (from 0·86 to 0·42) in the value of this constant over the range 0° to −18° C., the general tendency being for lower values to be obtained for the higher wet-bulb depressions.

A new theoretical treatment has recently been published by Robitzsch [67], in which it is shown that when the wet bulb is water-covered, the constant should decrease from a mean value of 0·505 at a wet-bulb temperature of 30° C. to a value of 0·477 at −20° C., and when ice-covered the value should vary from 0·429 at 0° C. to 0·432 at −60° C. wet-bulb. As at low temperatures the value obtained for the relative humidity becomes much more dependent on the value used for the psychrometric constant, there appears to be need for further investigation to fix more definitely the tables to be used with wet- and dry-bulb thermometers, before such instruments attain a high order of precision for humidity determinations in this region.

REFERENCES

(56) Dunmore, F. W. *Bur. Stand. J. Res., Wash.*, **20**, 723 (1938).
(57) Ebert, H. *Die Physik*, **6**, 47 (1938).
(58) Ebert, H. *Z. ges. Kälteindustr.* **44**, 127 (1937).
(59) Griffiths, E. and Awbery, J. H. *Seventh Int. Congr. Refrig.* (1936).
(60) Moss, E. B. *Proc. Phys. Soc.* **46**, 450 (1934).
(61) Hixson, A W. and White, G. E. *Industr. Engng. Chem. Analytical Edn.* **10**, 235 (1938).
(62) Winkler, C. A. *Canad. J. Res.* **17**, Section D, 35 (1939).
(63) Stack, S. S. *Gen. Elect. Rev.* **41**, 106 (1938).
(64) *Aspirations Psychrometer Tafeln*, 2nd Edn. (Vieweg und Sohn, Braunschweig, 1927.)
(65) Whipple, F. J. W. *Proc. Phys. Soc.* **45**, 307 (1933).
(66) Awbery, J. H. and Griffiths, E. *Proc. Phys. Soc.* **47**, 684 (1935).
(67) Robitzsch, M. *Met. Z.* **55**, 418 (1938).

§ 5. HEAT TRANSFER TO BOILING LIQUIDS

The rate at which heat is transferred from a hot surface to a liquid with which it is in contact is a factor of considerable practical importance. It exerts control in such diverse fields as those of the quenching of steel ingots, the commercial evaporator, and even that of the domestic boiler, yet only during the past few years has attention been directed to the physical nature of this common process. About fifty years ago Lang [68] noted that a maximum overall rate of heat transfer to a brine evaporator was obtained at a certain critical value of between 40 and 50° C. for the temperature difference, ΔT, between the liquid and heating surface, but apparently he did not investigate the matter further. A maximum rate of heat transfer corresponding to about 74,000 B.T.U. per sq. ft. per hour appears to have been just reached by Jakob and Linke [69] for the heat transfer from a horizontal surface to water. In a subsequent description of these experiments Jakob [70] states: "In this region the number of steam bubbles formed on the plate becomes so large that the bubbles disturb one another as to their stirring effects. We even observed a drop in α (the heat transfer coefficient in the equation $H = \alpha \,.\, \Delta T$) when increasing the heating energy still more. This of course was connected with a rapid rise of temperature, soon leading to a burning through of the heating resistance." In this experiment ΔT was only of the order of 15° C. This work, which dealt with the part played by bubbles of vapour in the transfer of heat by a fluid, was mentioned in Volume IV of these *Reports*, and was the first systematically planned research into the process of boiling. The investigation however appears to have been dropped at this interesting stage, and it was Nukiyama [71] who first conclusively demonstrated the existence of a definite maximum in the rate of heat transfer to a boiling liquid and realized its significance since it occurs at comparatively low values of ΔT. Nukiyama made measurements on an electrically heated wire immersed in water in a horizontal position. At first his results agreed with Jakob's in that α increased approximately linearly with ΔT, the increase being accompanied by the formation of more and more bubbles. In an experiment with a platinum wire 0·14 mm. in diameter, α reached a maximum value of nearly 20 cal. per cm^2 sec. when the value of ΔT was between 40° and 50° C., but when ΔT was increased to 150° C., α fell to 12·5 cal. per cm^2 sec. and then increased relatively slowly with further increase in ΔT. The sudden decrease from the maximum value appears to be associated with the formation of a film of vapour which completely envelops the hot surface. The heat transfer through this film occurs mainly by conduction and radiation, and relatively few bubbles escape from the film. It has been recognized by Scott [72] that the presence of an enveloping vapour film of this type accounts for the initial rate of cooling during quenching being slower than that which results after the ingot has cooled sufficiently to allow the quenching fluid to make direct contact with the metal.

The subject has since been treated in America at the hands of the following, who all obtain results in general agreement with that of Nukiyama:—Drew and Mueller [73]; Sauer, Cooper, Akin and McAdams [74]; and Rhodes and

Bridges [75]. Drew and Mueller gave a comprehensive account of the subject, and referred to the two types of boiling as nucleate and film boiling. They made observations on the overall heat transfer from a steam-heated copper coil immersed in various organic liquids, and found the transition from nucleate to film boiling to occur at temperature differences of the order of 30 to 50° C. but to vary with the nature of the liquid. They drew attention to the similarity of the present phenomenon of film boiling with that of the spheroidal state and gave a résumé of the somewhat scattered literature on this subject. Sauer and his collaborators gave results for several different metal surfaces and various fluids, including benzene and methyl alcohol. They found that the maximum attainable heat flux differed with the different surfaces, that for iron being the greatest. Rhodes and Bridges employed a technique which avoided the use of thermocouples to measure the surface temperature. Hot mercury was circulated through a short length of tubing immersed in the liquid under test, and the rate of flow of the mercury, and its temperature immediately before and after the boiling chamber, were observed. The rate of heat transfer from the mercury to the inner wall of the pipe was then calculated from the Dittus-Boelter equation. The conductance of the tube wall could also be calculated, and, having measured the total heat transfer, it was then possible to calculate the average resistance between the outer surface of the tube and the liquid. From the overall difference in temperature between the mercury and the liquid, the temperature difference between the outer surface and the liquid could then be evaluated. Using a steel tube and clean water, temperature differences of 60° C. were reached without the occurrence of film boiling, but, for the same tube coated with a film of wax, film boiling was established at a temperature difference of only 45° C. In this latter experiment the tube was seen to be surrounded by a continuous layer of vapour, and the heat transfer fell to about one-tenth of the value obtained with nucleate boiling and a temperature difference of only 20° C. A somewhat similar effect was produced when a small quantity of mineral oil was smeared over the tube, and in this instance the heat transfer was suddenly increased when a little sodium chloride was added to the water. A chromium-plated tube was found to give the transition to film boiling at a lower temperature difference, presumably owing to the lack of wettability of its surface. An explanation of these experiments can be found in the earlier observations of Fritz [76] on the growth of bubbles. The bubbles which formed under water on a metal having an oily surface which was not easily wetted by the water were observed to be attached to the surface by a broad base, whereas those which formed on a clean, easily wetted surface merely made point contact with the surface. In the former case the conditions are obviously more suitable for the establishment of an enveloping film of vapour. There appears to be need for further investigations in this field before the essential factors are fully understood and can be correlated as they have been for other types of convective heat interchange.

REFERENCES

(68) LANG, C. *Trans. Inst. Engrs. Shipbuilders Scot.* **32**, 279 (1888).
(69) JAKOB, M. and LINKE, W. *Phys. Z.* **36**, 267 (1935).
(70) JAKOB, M. *Mech. Engng., N.Y.*, **58**, 643 (1936).
(71) NUKIYAMA, S. *J. Soc. Mech. Engrs., Japan*, **37**, 367, S 53 (1934).
(72) SCOTT, H. *Trans. Amer. Soc. Metals*, **22**, 577 (1934).
(73) DREW, T. B. and MUELLER, A. C. *Trans. Amer. Inst. Chem. Engrs.* **33**, 449 (1937).
(74) SAUER, E. T., COOPER, H. B. H., AKIN, G. A. and MCADAMS, W. H. *Mech. Engng., N.Y.*, **60**, 669 (1938).
(75) RHODES, F. H. and BRIDGES, C. H. *Industr. Engng. Chem.* **30**, 1401 (1938).
(76) FRITZ, W. *Phys. Z.* **36**, 379 (1935).

§ 6. CONDENSATION, EVAPORATION AND CONVECTION

Just as in the previous section we have seen that the rate of heat transfer from a metal surface to a boiling liquid depends to a large degree on the wettability of the surface, so does this factor considerably influence the heat transfer due to the condensation of a vapour such as steam. If the cooling surface is wettable the condensation takes place as a continuous film of liquid, whereas drop-wise condensation occurs on a non-wettable cooling surface. It is found that the heat transfer in the latter case may be ten or twenty times as great as for film condensation, and it is only since visual observation has established the existence of the two types of condensation that an explanation has been forthcoming for some of the remarkable discrepancies observed in the results of earlier workers and for sudden changes in the performance of condensation plant.

A review of this interesting subject has recently been published by Dr Margaret Fishenden [77], so reference will only be made to the work of Gnam [78], which escaped notice in this treatment. Gnam studied the condensation of water vapour on the external polished surface of a chromium-plated brass tube, which was water-cooled internally. Drop-wise condensation was observed after treatment of the tube surface with liquids of low surface tension, such as lubricating oil, soft soap, or petroleum, and film condensation after treatment with caustic potash solution, which has a higher surface tension than water. Thermocouples attached to the tube revealed that a local fluctuation in tube temperature accompanied the formation of each drop, and it was estimated that corresponding fluctuations of the order of 25 per cent resulted in the heat transfer. The measurements confirmed that the heat of condensation was much greater for the drop-wise condensation, but since the resistance between the internal wall and the water was large it did not follow that the heat transfer from the external vapour to the internal water stream was increased in like proportion.

There have been several determinations of the evaporation of water from surfaces since this subject was dealt with in vol. IV of these *Reports*. Millar [79] has published the results of some wind-tunnel experiments carried out on water retained in rectangular pans measuring $25{\cdot}4 \times 50{\cdot}8$ cm. and $12{\cdot}0 \times 11{\cdot}2$ cm., set flush with the floor of the tunnel, at air velocities ranging from 89 to 890 cm. per sec. At any one air velocity, u, the rate of evaporation is found to be proportional to the difference in the vapour concentration. Empirical relations

are derived which show the rates from the large and small pans to be proportional to $u^{0.77}$ and $u^{0.725}$ respectively. It is concluded, however, that such formulae have only limited application, and the theory of turbulent mass exchange is therefore used to develop equations for the evaporation from both finite and infinite free water surfaces. The formulae obtained are in agreement with the principle of dynamic similarity. After the results for the larger pan have been used to derive a value for a constant which occurs in the equations, the theory succeeds quite well in predicting the rate of evaporation from the smaller pan. It is also applied with fair success to Rowher's data [80] for a 1-m. tank and an 85-ft. diameter reservoir.

Examination of other recent data in the light of Millar's theory reveals the results of Lurie and Michailoff [81] to be some 20 per cent greater than predicted, whereas data obtained by Shepherd, Hadlock and Brewer [82] are high to a still greater extent. In the last-mentioned experiments, the water was contained in a pan 1 ft. square and 2 inches deep, which was mounted in the centre of a wind tunnel. It would seem that the high rate of evaporation may have resulted from the increased turbulence set up in the air stream by the 2-inch vertical edge of the pan.

Much less attention has been devoted to the case in which the air is incident normally on the evaporating surface. Reference is usually made to some results reported by Carrier [83] which indicated the rate of evaporation for this case to be about twice as great as when an air stream of the same velocity passed tangentially over the surface. No details were given in Carrier's paper regarding the dimensions of the surface for which this result was obtained. Probably the surface was quite small, since the information was apparently intended to be applied to the case of a wet-bulb thermometer, and, in the light of recent work, the acceptance of this result as being of general application is unsound, since the size and shape of the surface as well as the nature of the air stream must exert a considerable influence. Kamei, Mizuno and Shiomi [84] have studied the rate of drying of slabs of clay 2 cm. in thickness which were coated with tin-foil so as to allow one face only to be exposed to the air stream. This exposed face measured 8 cm. wide by 5 cm. long, and when it was saturated with water they found the rates of drying for the facing, tangential and reversed positions to be in the ratio 1·0 : 0·92 : 0·76. Kamei [85] has since published a more detailed account of this investigation into the drying of various solid materials, a subject which has also been dealt with by Gilliland [86].

During the past year Molstad, Farevaag and Farrell [87] have published results for the rates of evaporation from free water surfaces ranging in diameter from 7·5 to 21·0 cm. and exposed to a perpendicular air stream, and here again the rates of evaporation do not differ greatly from those which would be expected for a tangential stream under similar conditions.

Sharpley and Boulter [88] have made some measurements of the evaporation of water from a 1-ft. diameter surface into quiet air, in which an ingenious optical-interference method is used to measure the rate of change of water level in the pan. In this work an attempt is made to correlate the rate of evaporation from the surface with the heat transfer due to natural convection.

In the presence of a forced air flow, the differential equations for the transfer of mass and heat are similar, and it is reasonable to expect that it should be possible to derive results for one process from those for the other. The relation between the two processes has been discussed by numerous workers. Schmidt [89] from a dimensional analysis concluded that

$$\alpha/\beta C_p = a/k = (\lambda/C_p\rho)/k,$$

where α and β are the coefficients of heat transfer by convection and mass transfer by evaporation in the equations

$$Q = \alpha A(T_{\text{surface}} - T_{\text{air}}), \text{ and } M = \beta A(C_{\text{surface}} - C_{\text{air}}),$$

λ, C_p and ρ are respectively the thermal conductivity, specific heat, and density of the air, and k is the coefficient of diffusion of water vapour into air. Experimentally, Hilpert [90] obtained a value of 0·92 for the ratio $\alpha/\beta C_p$. which is in good agreement with the theoretical value for water evaporating into air. Kirschbaum [91] has shown how the Mollier diagram can be employed as a means of deriving a relation between mass- and heat-transfer data. This treatment allows for the influence of heat transferred by direct radiation, and the conclusions are shown to be admirably supported by Hilpert's experimental data.

Chilton and Colburn [92] have suggested a semi-empirical dimensionless method of correlating heat- and mass-transfer data which has since been used by chemical engineers, and yet another treatment of the problem has been given by Ackermann [93]. Chilton and Colburn's method rests on the assumption that the "heat-transfer factor" $(h/cG)(c\mu/k)^{2/3}$ and the corresponding "mass-transfer factor" $Kp_{gf}(G/M)^{-1}(\mu/\rho k_d)^{2/3}$ are identical functions of the Reynolds number DG/μ. In these expressions, h is the film coefficient of heat transfer, D a linear dimension, c the specific heat at constant pressure, G the mass velocity, μ the viscosity, k the thermal conductivity, K the molar mass-transfer coefficient, p_{gf} the logarithmic mean partial pressure of the inert component of the film, M the mean molecular weight, ρ the density, and k_d the diffusion coefficient, any self-consistent set of units being used. On the basis of McAdams' correlation of existing data for the heat transfer from cylinders to a transverse air stream, Chilton and Colburn have plotted the best mean curve for $j = (h/cG)(c\mu/k)^{2/3}$ against DG/μ, and they suggest that this curve can be used for deriving mass-transfer data. Another curve of this kind is given for use in connection with problems involving mass interchange between a plane surface and a tangential air stream.

For a concise review of the general subject of heat transmission, reference should be made to a paper by Drew, Hottel and McAdams [94]. As well as containing results for both forced and natural convection for the standard types of surfaces, there are sections on condensation and the heat transfer in boiling liquids, and an account of recent progress in the field of radiant-heat transmission.

REFERENCES

(77) Fishenden, Margaret. *Engineering*, **145**, 643 (1938).
(78) Gnam, E. *ForschArb. IngWes.* **8**, 17 (1937).
(79) Millar, F. G. *Canad. Met. Mem.* **1**, no. 2, 41 (1937).
(80) Rowher, C. *Tech. Bull. U.S. Dep. Agric.* no. 271 (1931).
(81) Lurie, M. and Michailoff, N. *Industr. Engng. Chem.* **28**, 345 (1936).
(82) Shepherd, C. B., Hadlock, C. and Brewer, R. C. *Indust. Engng. Chem.* **30**, 388 (1938).
(83) Carrier, W. H *J. Amer. Soc. Heat. Vent. Engrs.* **24**, 240 (1918).
(84) Kamei, S., Mizuno, S. and Shiomi, S. *J. Soc. Chem. Ind., Japan*, **460–3** B, 38 (1935).
(85) Kamei, S. *Kyoto Coll. Eng. Mem.* **10**, 65 (1937).
(86) Gilliland, E. *Industr. Engng. Chem.* **30**, 506 (1938).
(87) Molstad, M. C., Farevaag, P. and Farrell, J. A. *Industr. Engng. Chem.* **30**, 1131 (1938).
(88) Sharpley, B. F. and Boelter, L. M. K. *Industr. Engng. Chem.* **30**, 1125 (1938).
(89) Schmidt, E. *Gesundheitsing.* **52**, 525 (1929).
(90) Hilpert, R. *ForschArb. IngWes.* no. 355 (1932).
(91) Kirschbaum, E. *ForschArb. IngWes.* **7**, 109 (1936).
(92) Chilton, T. H. and Colburn, A. P. *Industr. Engng. Chem.* **26**, 1183 (1934).
(93) Ackermann, G. *ForschArb. IngWes.* **8**, 1 (1937).
(94) Drew, T. B., Hottel, H. C. and McAdams, W. H. *Trans. Amer. Inst. Chem. Engrs.* **32**, 271 (1936).

§ 7. RADIATION

Reference should be made to the paper by Drew, Hottel and McAdams [94] for an account of the progress made from 1930 to 1936. Some more recent contributions are mentioned below.

The amount of heat radiated from a surface varies directly with its emissivity, and reaches a maximum value for a black body (emissivity=1). When modifications are considered with a view to either increasing or decreasing the total heat interchange between a surface and its surroundings, it is well to remember this easily varied factor, rather than to consider only the means whereby the convective transfer can be modified. Rubert [95] has published the results of an experimental study of radiators of the type commonly installed for steam or hot-water heating. He obtained emissivity values of 0·88, 0·82, 0·81, 0·40 and 0·28 respectively for the following types of surface: black paint, clean cast iron, black scale, aluminium paint, and bright steel. For the particular conditions of Rubert's tests (radiator temperature 215° F., air temperature 70° F.) the heat loss by convection was invariably greater than that by radiation. In the case of the clean cast iron surface, the loss by radiation amounted to 43 per cent of the total, and after coating with aluminium paint this figure was reduced to about 28 per cent.

A more exact series of measurements has been carried out by McDermott [96]. He has published data for the total thermal emissivities of various sheet materials and has compared the results with existing values. He has endeavoured to minimize errors arising from direct reflection and obtains somewhat lower values for some of the materials of low emissivity. A value of only 2·8 per cent is obtained for bright aluminium foil, whereas McAdam's *Heat Transmission* gives 3·9 to 5·7 for polished aluminium.

Wilson has also published the results of a number of tests which give data showing the effect of surface treatment on heat transfer. He was primarily concerned with the loss of heat from electrical equipment at only moderate temperatures above the surroundings, and chose, in expressing his results, to combine the convective and radiation transfers and to assume Newton's law to hold, $H=eA(T-T_0)$. The constant e now relates to the net heat transfer, and in experiments on a copper surface it was found that coatings of shellac varnish and oil paints gave the highest values of e, values about 40 per cent in excess of that for the natural copper being obtained.

Kalous [98] has dealt with the theoretical aspects of radiant heating of rooms, and Bennett and Hartley [99] have described an apparatus for measuring the radiation from domestic heaters. The heater is mounted on a central platform, about which a receiving thermopile carried on a semicircular arc moves over the surface of a hemisphere. The transit, which occupies about 40 seconds, is made with the receiver in various positions so disposed to give alternately positive and negative deflections of the galvanometer to which it is connected. Once the calibration factor of the instrument has been established, the total quantity of heat radiated by the appliance under test can be determined.

Gerlach and Buhl [100] have described an instrument for recording the total amount of radiation received at a given place in a given time. In this instrument the radiation is allowed to fall on the vanes of a Crookes radiometer. A very small mirror is attached to the rotating system and is so adjusted that, at each rotation of the vanes, a flash of reflected light falls on a photo-cell. The current impulse produced in this way is amplified and operates a counter. Tests have shown that the speed of rotation is approximately proportional to the incident radiation, so that the number shown by the counter is a measure of the quantity of radiation which has been received.

A very complete account of the absolute measurement of the intensity of thermal radiation has been published by Guild [101]. It is learnt from this paper that Professor Callendar used to test British radiometric instruments, and that when he died, the equipment was sent to the National Physical Laboratory. Since 1920 the scale of thermal radiation has been maintained at that laboratory by means of a Callendar radio-balance of improved design. This instrument consists of two massive copper discs, thermally insulated from their supports and connected by a thermocouple which indicates any difference in their temperatures. Arrangements are made for one of the discs to be heated either by the radiation or by the generation of an electric current in a grid attached to its blackened front surface. Following a preliminary heating which brings the system into a suitable condition for the experiment, a measurement is made by heating the front of the disc alternately by the radiation under test and by a current which develops approximately the same amount of heat. Each period of heating lasts exactly 100 seconds, and the heating is applied after a cooling interval of 100 seconds. The temperature difference of the discs at the beginning and end of each 100-second interval is observed on the scale of a galvanometer. To the rise

in temperature for each heating is added half the sum of the preceding and following falls and the mean value of this quantity obtained for the radiant and electrical periods of heating gives the ratio of the intensities for the two systems of heating. A series of comparison experiments has been carried out which shows that the results of this new instrument agree with those of Callendar's two drift-radiometers to about 1 part in 1000. The N.P.L. scale of thermal radiation has also been compared with all other scales used for the calibration of industrial and meteorological radiometers. It agreed with the Callendar scale to within the accuracy of comparison, estimated at $\pm 0{\cdot}5$ per cent, and with the National Bureau of Standards scale to $\pm 0{\cdot}2$ per cent. The Smithsonian scale of 1913 was found to be 2·3 per cent too high (this agrees to 1 in 1000 with the error of 2·3 per cent already noted at that Institution), and the Ångström scale was considered to be about 0·5 per cent too low.

In the industrial world the drying of lacquer finishes and so forth is being expedited by the use of near infra-red radiation. Ickis and Haynes [102] have described an installation consisting of a drying tunnel with the walls lined with banks of 250-watt lamps housed in gold-plated reflectors which can be used for drying articles of various sizes up to that of a complete motor-car body. Curves are given showing the relative output from various sources of infra-red energy, and the transmission of infra-red energy by water and lacquer films of various thicknesses. A drying lamp running at a colour temperature of about 2500° K. is considered best for removing moisture from pulpy substances, since radiation corresponding to 1000° K. will mostly be absorbed in the first millimetre of water, whereas an increase in temperature gives a radiation which penetrates more deeply into the body of the water before being absorbed.

REFERENCES

(95) Rubert, K. F. *Bull. Cornell Univ. Engng. Experiment Sta.* no. 24, July 1937.
(96) McDermott, P. F. *Rev. Sci. Instrum.* **8**, 185 (1937).
(97) Wilson, W. *Beama J.* **43**, 91 (1938).
(98) Kalous, K. *ForschArb. IngWes.* **8**, 170 (1937).
(99) *Engineering*, **145**, 717 (1938).
(100) Gerlach, W. and Buhl, O. *Phys. Z.* **38**, 795 (1937).
(101) Guild, J. *Proc. Roy. Soc.* A, **161**, 1 (1937).
(102) Ickis, L. S., Jr., and Haynes, H. *Gen. Elect. Rev.* **42**, 145 (1939).

§ 8. SPECIFIC AND LATENT HEAT DETERMINATIONS

In recent volumes of these *Reports* references have been made to the increased importance which has become attached to accurate specific-heat determinations, since they furnish a means of testing the various theories of order-disorder transformation.* Sykes and his collaborators, who have been to the fore in conducting these experimental investigations, have now applied their technique in an endeavour to test experimentally the various theories of age hardening [103]. Swindells and Sykes [104] have published

* Some account of these theories will be found in the article on p. 369 of this volume. [Editor.]

specific heat *vs.* temperature curves for five typical age-hardening alloys, and have discussed these results in the light of hardness measurements. In the case of silver-copper and copper-beryllium alloys they conclude that the maximum hardness is attained when most of the chemical energy associated with the supersaturated solid solution has been evolved. This is consistent with precipitation having taken place. With duralumin and aluminium-copper alloys, on the other hand, they find that the maximum hardness occurs before precipitation takes place. In these cases the hardening is more probably due to the formation of clusters or groups of atoms inside the parent lattice than to precipitation. The reasons for this varying behaviour have not yet been ascertained, but it would certainly seem that specific-heat measurements are likely to lead to a better understanding of the phenomenon.

In vol. IV of these *Reports* (p. 85) existing data for the specific heat of nickel over the temperature range 0° to 700° C. were tabulated. Sykes and Wilkinson [105] have now determined the specific heats of four samples of nickel up to about 600° C., and have critically reviewed the results of previous investigators. In this way, they derived values for the most probable specific heat of nickel, which they consider to be correct to within 0·5 per cent up to the Curie temperature and to 1 per cent at higher temperatures. These new results have been examined in connexion with the theory proposed by Stoner [106], which suggested that the specific heat at constant volume of a ferromagnetic metal could be represented as the sum of three terms involving the lattice, electronic, and magnetic contributions to the specific heat, and the problem now appears to call for a rather more detailed treatment than that proposed originally. Another determination of the specific heat of nickel has been carried out by Néel [107]. The values obtained above the Curie point are about 3·5 per cent lower than those of Sykes and Wilkinson. In Néel's method an ingot weighing about 100 gm. was supported in an evacuated enclosure and electrically heated through small temperature intervals of about 2° C. The heat lost by radiation was determined from observations of the rise in temperature of a jacket of known heat capacity which completely surrounded the specimen.

Iron is another metal for which more than usual interest has been shown in the variation of its physical properties with temperature. Naeser [108] had carried out water-calorimetric measurements which were reported as showing that the specific heat *vs.* temperature curve of iron had no less than six sharp discontinuities between room temperature and the Curie point. This surprising result was supported to some extent by the specific-heat data of Dearden [109] and by several earlier measurements of other properties. During the past year Sykes and Evans [110] have endeavoured to clarify the position, and have determined the specific heat *vs.* temperature curves for four different samples of iron during heating in their adiabatically controlled furnace. The curves obtained in this way over the range 70° to 450° C. are identical to ± 1 per cent for the four irons and are free from discontinuities. Following a careful examination of existing data the most probable values for the specific heats of iron at temperatures of 100°, 200°, 300° and 400° C. are

considered to be 0·114, 0·126, 0·137 and 0·146 g.cal. per gm. °C. The paper devotes considerable space to the question of the existence of discontinuities, and it is concluded that those found in Naeser's work must either have resulted from experimental error, or from the effects of hydrogen or other dissolved gases. The discussion of the paper was unusually interesting and served to clarify the matter further. Dearden wrote that his results were based on only one series of measurements, which the present work had convinced him must have been in error. Moser wrote, saying he had tested Naeser's actual specimen by a method similar to that used by Sykes, and had failed to find the discontinuities reported by Naeser in the ranges 90° to 135° C. and 180° to 220° C. He thought Naeser's results must be due to residual gases in the iron. Yet, as Sykes remarked, Naeser had worked with vacuum-melted electrolytic iron!

At about the same time the results of an experiment, which seems to confirm that dissolved gases can cause an effect of this kind, were published by Zuithoff [111]. He made measurements of the mean specific heat of pure electrolytically precipitated iron contained in platinum vacuum crucibles when dropped into a metal block calorimeter. At first he obtained a number of discontinuities in the derived specific heat *vs.* temperature curve, which corresponded roughly with those observed by Naeser, and in these experiments the crucible, after cooling, was observed to be somewhat inflated, as though gases had been evolved from the heated metal. After repeatedly heating to 950° C. the inflation was no longer observed, and the specific heat *vs.* temperature curve was then smooth apart from a hump in the neighbourhood of 155° C.

A small hump in this region does appear to be characteristic of carbon steels, for it has been observed both by Sykes and Evans and by Awbery and Snow [112]. Awbery and Snow have measured the specific heats of eight carbon steels up to about 950° C. Their method somewhat resembles that of Sykes but has an important difference in that the outer furnace consists merely of a hollow box-like enclosure formed of kanthal ribbon stretched across a light insulating framework. This construction ensures that the furnace is of very low heat capacity and enables its temperature to be readily varied so as to maintain the specimen always at the same temperature as that of a thin-walled steel enclosure interposed between the specimen and the furnace. This arrangement has worked very well and has enabled heat interchange by radiation to be prevented even during the arrest points of the transformation region. It has thus been possible to obtain information regarding the heat quantities associated with the various transformation points of steels.

The specific-heat data obtained over the lower temperature range for four of the steels are of particular interest, in that these steels appear to be obtained from the same source as the steels for which determinations were made by Sykes and Evans up to 450° C. The two sets of data are seen to be in close agreement, certainly to within about 2 per cent, which is the order of accuracy to which the values can be read off from the published curves. In view of the number of determinations which have now been carried out in Sykes' apparatus, this independent confirmation of the accuracy of his results is very satisfactory.

The above steels were tested when in the annealed condition; Sykes and Evans obtained appreciably different results after the steels had been quenched, the differences being attributed to heat evolved during decomposition of austenite.

Manganese would appear to be a metal for which interesting data could be obtained in an apparatus of this type, for it has been reported as having a number of transformations in the solid state. Gayler [113] mentions four, at temperatures of 682°, 742°, 1024° and 1191° C., whilst Yoshisaki [114] mentions three, at 706°, 1087° and 1150°.

A new determination has been made of the specific heat of bismuth over the range −80° to 120° C. Goetz and Jacobs [115] had reported the existence of discontinuities, or at least rapid changes, in the expansion coefficient of the bismuth lattice, and Bronson and MacHattie [116] considered that similar peculiarities might be expected in the heat capacity, if Grüneisen's law were obeyed. No such deviations were found, the specific heat in joules per gm.°C. being given by the equation $C_p = 0{\cdot}1238_5 + 4{\cdot}45 \times 10^{-5}\, t$, over the range of their experiments.

A re-determination of the latent heat of fusion of aluminium has been carried out by Awbery [117] using an electric-heating method. The specimen was contained in a sintered alumina container and could be either heated or cooled within a furnace. A previous calibration had been made of the heat transfer between the furnace and the container. Three melts gave a mean latent heat of 93·9 cal. per gm. and three freezes a mean of 88·3. The determination was admittedly not of the greatest accuracy, but, as existing data for the latent heat could be divided into two groups of values ranging from 64 to 82 and 92 to 94 cal. per gm., the new result definitely confirms the higher group of values. The latent heat of fusion of aluminium is considered by Awbery to be $92{\cdot}5 \pm 0{\cdot}4$ cal. per gm.

With some substances deviations from the normal specific-heat curve occur shortly before the melting-point is reached. Carpenter and Harle [118] have reported that in the case of iodine the specific heat passes through a flat maximum, and then decreases slightly on approaching the melting point. For the alkali elements [119], [120] on the other hand, there appears to be an abnormal rise of specific heat extending over a considerable range of temperature below the melting point. This abnormality is thought to be due to a gradual breakdown of the lattice, as the X-ray diffraction patterns also disappear well below the melting point. Ubbelohde [121] has observed rather peculiar results for some of the long-chain compounds, notably for octadecane. The apparent specific heat rises to a sharp maximum some 15° C. below the normal setting temperature, as determined on the same sample. This is considered to arise from the occurrence of pre-melting, and, since it is observed in samples containing less than 1 per cent of neighbouring paraffins as impurity, it cannot be attributed to the presence of impurities, and is thought to be due to a process taking place within the solid phase as the melting point is approached. The existence of effects of this type makes the evaluation of the latent heat of fusion somewhat arbitrary.

There has been considerable activity in the determination of the essential thermophysical properties of modern refrigerants to low temperatures. The following summarizes the results obtained by Riedel [122] for trifluor-trichlor-ethane (Freon 113); v', the specific volume of the liquid from $-30°$ to 80° c., is given by

$$v' = 0{\cdot}6169(1 + 1{\cdot}390 \times 10^{-3}t + 2{\cdot}74 \times 10^{-6}t^2 + 1{\cdot}22 \times 10^{-8}t^3);$$

p, the vapour pressure, in atmospheres absolute, by

$$\log p = 9{\cdot}6842 - 2099/T - 1{\cdot}3505 \times 10^{-2}T + 0{\cdot}1171 \times 10^{-4}T^2.$$

By means of the Clausius-Clapeyron equation the latent heat of vaporisation r, in k.cal. per k.gm., is derived as

$$r = 37{\cdot}90 - 6{\cdot}75 \times 10^{2}t - 0{\cdot}7 \times 10^{-4}t^2,$$

whilst the specific heat of the liquid from $-30°$ to 61° c. is given by

$$C = 0{\cdot}2172 + 0{\cdot}031 \times 10^{-2}t + 0{\cdot}012 \times 10^{-4}t^2.$$

Riedel [123] has since published the results of a similar determination on pure sulphur dioxide over the range $-60°$ to 30° c., and the data are in reasonably good agreement with another recent determination by Giauque and Stephenson [124]. Plank and Seger [125] have obtained the necessary data to enable the Mollier diagram for difluor-dichlor-methane to be extended down to a temperature of $-80°$ c.

A paper of interest to the users of combustion calorimeters has been published by Huffman [126]. It appears that the First Report of the Commission Permanente de Thermochemie de l'Union Internationale de Chimie (Paris, 1934) accepted benzoic acid as a standard material for use in bomb calorimetry. At a later meeting of this Commission (1936) succinic acid was formally adopted as a secondary standard, and a preliminary value was given for its isothermal heat of combustion. In view of the fact that a final value has not yet been accepted, the value of 30231 cal. per gm. is now suggested for standard conditions at 25° c. For hippuric acid crystallized under water, the value of 5628·6 cal. per gm. is obtained and the suggestion is made that hippuric acid might usefully be adopted as a secondary standard for investigations on nitrogen-containing compounds.

REFERENCES

(103) Gayler, M. *J. Inst. Met.* **60**, 1, 249 (1937).
(104) Swindells, N. and Sykes, C. *Proc. Roy. Soc.* A, **168**, 237 (1938).
(105) Sykes, C. and Wilkinson, H. *Proc. Phys. Soc.* **50**, 834 (1938).
(106) Stoner, E. C. *Phil. Mag.* **22**, 81 (1936).
(107) Néel, L. *C.R. Acad. Sci., Paris*, **207**, 1384 (1938).
(108) Naeser, G. *Mitt. K.-Wilh.-Inst. Eisenforsch.* **17**, 185 (1935).
(109) Dearden. *Iron & Steel Inst. Carnegie Scholarship Memoirs*, 17 (1928).
(110) Sykes, C. and Evans, H. *J. Iron Steel Inst.* **138**, 1258 (1938).
(111) Zuithoff, A. J. *Proc. Acad. Sci. Amst.* **41**, 265 (1938).
(112) Awbery, J. H. and Snow, A. *Iron Steel Inst.* (Special Report no. 24); *Second Report of the Alloy Steels Research Committee*, 216 (1939).
(113) Gayler, M. *Nature, Lond.*, **124**, 840 (1929).
(114) Yoshisaki, H. *Sci. Rep. Tôhuko Univ.* **26**, 182 (1937).

(115) JACOBS, R. B. and GOETZ, A. *Phys. Rev.* **51**, 159 (1937).
(116) BRONSON, H. L. and MACHATTIE, L. Æ. *Canad. J. Res.* **16**, 177 (1938).
(117) AWBERY, J. H. *Phil. Mag.* **26**, 776 (1938).
(118) CARPENTER, L. G. and HARLE, T. F. *Phil. Mag.* **23**, 193 (1937).
(119) BIDWELL, C. C. *Phys. Rev.* **28**, 584 (1926).
(120) CARPENTER, L. G., HARLE, T. F. and STEWARD, C. J. *Nature, Lond.*, **141**, 1015 (1938); see also *Phil. Mag.* **27**, 551 (1939).
(121) UBBELOHDE, A. R. *Trans. Faraday Soc.* **34**, 282 (1938).
(122) RIEDEL, L. *Z. ges. Kälteindustr.* **45**, 221 (1938).
(123) RIEDEL, L. *Z. ges. Kälteindustr.* **46**, 22 (1939).
(124) GIAUQUE, W. F. and STEPHENSON, C. C. *J. Amer. Chem. Soc.* **60**, 1389 (1938).
(125) PLANK, R. and SEGER, G. *Z. ges. Kälteindustr.* **46**, 41 (1939).
(126) HUFFMAN, H. M. *J. Amer. Chem. Soc.* **60**, 1171 (1938).

§ 9. THE CONDUCTION OF HEAT

The general subject of the conduction of heat, and in particular the relation between this property and the conduction of electricity in electrical conductors, was given a moderately detailed treatment in last year's *Report*. The present account will accordingly be limited to thermal conductivity determinations for other than metallic conductors. Johnson [127] has determined the thermal conductivity, K, and electrical conductivity, σ, of eighteen light-weight fire-bricks for which K ranged from 0·000318 to 0·000496 g.cal. per cm. sec.°C.* and four building bricks for which K ranged from 0·00164 to 0·00183, and suggests a simple method of correlating these two properties. When either thermal or electrical conductivity was plotted against the porosity, as measured by the weight per cent of water absorbed by the brick, he obtained the same type of curve. It followed that a straight line was obtained when the ratio of the thermal to the electrical conductivity (which varied in these experiments from $15{\cdot}7\times10^8$ to $83{\cdot}6\times10^8$) was plotted against the percentage water absorption. This work suggests a simple means whereby an estimate can be formed of the thermal conductivity of a sample of brick.

Presumably the above determinations apply to dried samples, since the thermal conductivity of an insulating material is dependent on its water content to a considerable extent. This fact can become one of considerable importance, particularly in the insulation of cold-stores, where the low temperature encourages the condensation of moisture on the insulation. Meissner and Immler [128] have recently published the results of some tests which show that calculations of heat transfer through the insulation become very difficult under such conditions. They compared the variation over the range −10° to 30° C. of the thermal conductivity of expanded cork of high moisture content with that of the dry material. Sufficient thermocouples were fixed in the slab to enable measurements of the thermal conductivity to be made for the cold-face, mean, and hot-face sections of the slab. When dry, a linear increase was obtained from a value of $0{\cdot}000101_4$ at −10° C. to $0{\cdot}000130_5$ at 35° C., the points for all three sections of the slab falling on this line. In the case of the sample of high moisture content, the values for the three sections

* All thermal-conductivity data given in this section will be expressed in these gm. cal. $cm.^{-1}$ $sec.^{-1}$ $°C.^{-1}$ units.

coincide over the range from −10° C. to 10° C. as the conductivity increases from $0{\cdot}000119_4$ to $0{\cdot}000133_3$; at about 10° C., however, sharp discontinuities occur and the results for each section of the slab follow a different curve. That for the cold-face section falls to a value of $0{\cdot}000130_5$ at 13° C. and then rises linearly to $0{\cdot}000144_4$ at 35° C. The conductivity of the mean section falls to $0{\cdot}000126_4$ at 16° C. and then rises to $0{\cdot}000140_2$, whilst that for the hot-face section falls to 0·000111 at 18° C. and ultimately increases to $0{\cdot}000134_7$ at 35° C. These variations are attributed to the movement of moisture within the material as a result of capillary action, and presumably the effective moisture content at a particular locality will depend both on the applied temperature gradient and the previous temperature history of the sample. Prior to equilibrium being attained, the actual movement of the water also forms an additional means of heat transfer.

Further results for the thermal conductivities of frozen moist insulators have been published by Nagaoka, Watanabe and Yasiro [129].

Griffiths [130] has published a paper containing brief descriptions of various types of apparatus in use at the National Physical Laboratory. Conductivity values are given for several corks and light-weight concretes, and also the equivalent conductivity of a multiple partition insulation. This consisted of parallel sheets of asbestos paper coated with aluminium foil, the sheets being spaced 1 inch apart and built up to a thickness of 5 inches. In the case of this cellular type of insulation a certain amount of convection can occur and the heat transmission from face to face varies with position. In the tests reported, the equivalent conductivity for a cold-face temperature of 20° C. and hot-face of 30° C. was 0·00023 with the surfaces vertical, 0·00017 with the surfaces horizontal and hot face uppermost, 0·00019 with the surfaces horizontal and cold face uppermost, or with the surfaces inclined at 45° and hot face uppermost, and 0.00022 as last but with the cold face uppermost. Further data as to the uses and insulating values of various types of aluminium-foil insulation, together with comparative values for other materials, have been published by Schmidt [131].

An apparatus of the plate type has been described by Koch [132] for the determination of absolute thermal conductivities up to 800° C., results being given for a hard porcelain and for a kieselguhr brick. The specimen is in the form of a square slab of 30 cm. side and 3 to 5 cm. thickness, a central square section of 10 cm. side being isolated from the main slab so that the outer portion serves as a guard ring. Below the specimen is an electric plate heater extending over the whole area of the specimen, but in which the energy supplied to the isolated central section can be measured independently. This is backed by another plate heater which serves to prevent downward loss of heat and is separated from the hot plate by a slab of poorly conducting material. The whole of the upper surface of the test slab is either cooled by a water-flow unit, or heated electrically to a temperature below that of its lower surface. In this apparatus the conductivity data are obtained with relatively small temperature differences between the hot and cold surfaces of the sample. Temperature differences of from 14° to 34° C. were employed

in the case of porcelain, for which the conductivity varied linearly from 0·0034 at 0° C. to 0·00435 at 800° C., whilst in the case of the kieselguhr brick the temperature differences varied from 67° to 97° C. and the conductivities from 0·000214 at 0° C. to 0·000328 at 400° C. and 0·000495 at 800° C.

Descriptions of two other plate methods, one an absolute and the other a relative method, have been given by Fritz and Diemke [133] which have been used to measure the thermal conductivities of 31 samples of coal and 3 samples of coke at 30° C. The absolute method, which also involved a guarded hot plate, was applied to a sample 100 mm. in diameter and from 15 to 20 mm. in thickness, the surface contacts being improved by means of paraffin. An accuracy of ± 1 per cent is claimed. The conductivities of anthracite were found to vary from 0·005 to 0·0086, of gas coals from 0·0046 to 0·0067, and of coke from 0·0172 to 0·0335. In one instance in which the tests were extended over the temperature range of 20° to 100° C. the conductivity increased linearly from 0·0062 to 0·0073. These values for the thermal conductivity of coal are seen to be of the same order of magnitude as the value obtained by Powell and Schofield [134] for the thermal conductivity of an amorphous carbon consisting of 80 per cent of petroleum coke combined with 20 per cent of lamp-black. These workers were primarily concerned, however, with measurements at high temperatures. For this purpose they used a method similar to that suggested by Mendenhall, for materials which are conductors of electricity, and applied by Angell [135] to aluminium and nickel. In this method a steady current of electricity is passed through a long cylindrical bar of the material, and the thermal conductivity is obtained from a measurement of the rate of energy dissipation in the central portion of the bar, and the difference in temperature between the axis and surface. Whereas in Angell's experiments this temperature difference only appears to have been of the order of 1° C., differences as great as 300° C. were obtained in some of the experiments on carbon. Up to about 1400° C. a linear increase is obtained in the thermal conductivity of carbon. Further increase in temperature tends to graphitize the carbon, and the thermal conductivity increases more rapidly. Values obtained on cooling were then found to be on a higher curve. Results are also given for the thermal conductivity of Acheson graphite up to about 2700° C. From 2000° C. upwards, the sample tested had a fairly constant thermal conductivity of about 0·03. It is interesting to note that whereas at normal temperatures the Lorenz functions of carbon and graphite are respectively about 17 and 200 times as great as the value normally obtained for metals, the values decrease rapidly with increase in temperature, and at 1700° C. the Lorenz functions of both materials are only four times as great as those of metals, whilst at 2500° C. that of graphite is only double the typical metallic value.

Turning now to consider work in connexion with the conductivities of liquids and gases, it is seen that Bates in collaboration with others [133] has applied a plate method to determine the thermal conductivities of mixtures of methyl and ethyl alcohol with water over the temperature range from 0° to 70° C. The temperature coefficient is found to reduce to zero for 50/50

mixtures of methyl alcohol and water and for 52/48 mixtures of ethyl alcohol and water. The values obtained for the two alcohols at 20° C. are in agreement to within 2 per cent with the values given in the International Critical Tables. Barrett and Nettleton, who prepared this section of the tables, suggested the use of the following equation to obtain the conductivity of binary mixtures:

$$K \sinh (100\,\mu) = K_1 \sinh (p_1\mu) + K_2 \sinh (p_2\mu),$$

where p_1 and p_2 are the percentages by weight of constituents having thermal conductivities K_1 and K_2 and μ is a constant depending on the constituents and temperature. For aqueous solutions of methyl alcohol, ethyl alcohol, and glycerine, Bates and collaborators find μ to have values of 0·90, 0·94 and 0·65 respectively, and to be independent of the temperature.

For measurements over a range of temperatures, the hot-wire method has several advantages over the plate method, but suffers from the disadvantages of end corrections on account of longitudinal conduction if the wire is thick, and temperature discontinuities at its surface if the wire is thin. Sherratt and Griffiths [137] have avoided the difficulties associated with the use of a thick platinum wire by welding loops of wire towards its ends, which can be independently heated to the same temperature as the working section of the wire. In an apparatus of this type measurements of the thermal conductivities of air, carbon dioxide and dichlor-difluor-methane (Freon) have been made over the temperature range 0° to 300° C.

Eucken and Englert [138] have also used a modification of the hot-wire method for the determination of the thermal conductivities of some liquids and solidified gases. In these experiments the observations were made during the variable state, the conductivity being calculated from the rate of temperature rise of the central wire when carrying a steady current. This procedure has the advantage that it avoids complications due to the thermal resistance between the heated wire and the material under test. The results obtained include values of 0·00108 and 0·00128 for the thermal conductivity of solid carbon dioxide at temperatures of −78·5° and −103·9° C. respectively, and a value of 0·00239 for solid ammonia at the latter temperature. A more general account of the theory of this method has since been published by Fischer [139], who points out that Eucken and Englert's expressions only hold under certain conditions. An analogous treatment to that for the hot-wire method is given for the plate method

REFERENCES

(127) Johnson, J. S. *J. Amer. Ceramic Soc.* **21**, 79 (1938).
(128) Meissner, W. and Immler, R. *Wärme u. Kältetechn.* **40**, 129 (1938).
(129) Nagaoka, Z., Watanabe, A. and Yasiro, Y. *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, no. 824, 1034 (1938).
(130) Griffiths, E. *J. Sci. Instrum.* **15**, 117 (1938).
(131) Schmidt, E. *Z. ges. Kälteindustr.* **44**, 163 (1937).
(132) Koch, W. *Phys. Z.* **39**. 431 (1938).
(133) Fritz, W. and Diemke, H. *Feuerungstechnik*, **27** (5), 129 (1939); see also *Phys. Z.* **40**, 361 (1939).

(134) POWELL, R. W. and SCHOFIELD, F. H. *Proc. Phys. Soc.* **51**, 153 (1939).
(135) ANGELL, M. F. *Phys. Rev.* **33**, 421 (1911).
(136) BATES, O. K., HAZZARD, G. and PALMER, G. *Industr. Engng. Chem. Anal. Edn.* **10**, 314 (1938).
(137) SHERRATT, G. G. and GRIFFITHS, E. *Phil. Mag.* **27**, 68 (1939).
(138) EUCKEN, A and. ENGLERT, H. *Z. ges. Kälteindustr.* **45**, 109 (1938).
(139) FISCHER, J. *Ann. Phys., Lpz.*, **34**, 669 (1939).

§ 10. THERMAL EXPANSION

Mention of this subject was last made in vol. III of these *Reports*, when a few practical instances were given to show the variety of problems in which the effects of thermal expansion become matters of importance. Yet another such instance occurs when two different materials are bonded together as in the glass-to-metal seals of valves and vacuum tubes. For the seal to withstand changes of temperature it is necessary to select a glass and metal with nearly the same expansion coefficient. This subject has been dealt with by Hull, Burger and Navias [140], who have made a study of the thermal expansion of iron-nickel alloys and have succeeded in finding one to match a normal type of glass. A new glass is also described, which can be sealed directly to iron.

Apart from its practical importance, thermal expansion is, as Robertson and Ubbelohde [141] have pointed out, one of the most important thermodynamic properties of a crystal in relation to its structure, since it provides information on the lattice vibrations, and thereby on the binding forces. These workers have derived expansion coefficients over the range 90° to 290° K. by X-ray spectroscopy for crystals of rock salt, oxalic acid dihydrate and resorcinol, and from the observed variations in the coefficient for different directions, and the effect of replacing hydrogen by deuterium, considerable light has been thrown on the structure.

A general account of the use of expansion data for non-metallic crystals to give information on their structure and inter-atomic forces has been published by Megaw [142].

In recent years increasing use is being made of X-ray spectroscopy as an accurate and relatively simple method of measuring expansion coefficients; the lattice expansion obtained in this way agrees in many instances with measurements made on the bulk material. Owen and his collaborators have been very active in this field of research, and mention might be made of Owen and Yates' [143] results for a 99·98 per cent nickel. In passing through the temperature of the magnetic transformation, nickel retains its face-centred cubic lattice, but with the loss of magnetism the parameter of the lattice suddenly increases, so that the expansion coefficient has a maximum value at 370° C. After passing through this region, the coefficient decreases and later increases, the curve above 420° C. being a continuation of that below 300° C.

Marick [144] has carried out similar experiments on cobalt for which there is also a marked expansion of the lattice at the Curie point. Iron, on the other hand, undergoes a sudden contraction at the Curie point [145], [146].

Most metallurgical laboratories are equipped with some form of dilatometer, frequently of a differential type, since the change in the rate of expansion which occurs at a transformation point offers a ready means of obtaining information as to metallurgical changes of phase. Ackley and Walters [147] have described an instrument for such observations which automatically records a differential expansion *vs.* temperature curve, obtained photographically on a revolving drum, and a true temperature curve drawn by a pen. The dilatation coordinate is obtained by an optical lever system which amplifies the difference between the expansion of the specimen and of comparison pieces. A 500-fold amplification is obtained, and the instrument can be used over the range 25° to 1100° C. to measure differential length changes of from 0·00005 to 0·01 inch.

Several dilatometric studies of iron and steels have been made and, as a result of a recent investigation in which about 60 curves were obtained for iron of high purity, Wells, Ackley and Mehl [148] give a value of 909·5° ± 1·0° C. for the temperature of the α–γ transformation. The work of Rosenbohm [146] on iron suggests the presence of a sharp maximum in the expansion coefficient *vs.* temperature curve at about 140° C. The kink was detected in each of five photographic records obtained on a 41 mm. rod in an apparatus capable of measuring length changes of 0·0001 mm. For the 5° C. interval between 140° and 145° C. the increment in length was $0{\cdot}0029_2$ mm., whereas at 135° C. and 155° C. the corresponding increments were $0{\cdot}0027_3$ and $0{\cdot}0026_9$ mm. respectively. It is claimed that the existence of a similar maximum between 100° and 150° C. can be detected in the results of an earlier determination by Mauer and Schmidt [149], though their paper only gave values for the mean coefficient from 20° to 50°, 100°, 150° and 200° C. respectively. The question of the existence of this point is an interesting one worthy of independent confirmation, particularly in view of the claims already mentioned that the specific-heat curve also has a maximum value at approximately the same temperature.

It is impossible to refer to the many determinations which have been made of the expansion coefficients of metals. Hidnert [150] has been a busy worker in this field, and in a recent paper [151] results are given for four heat-resistant alloys of modern development, namely, Hoskins alloy No. 10 and Kanthal *A*, *A*1, and *D*. It was found that heat treatments at temperatures up to 1400° C. caused considerable grain growth, and permanent increases in length of from 1 to 2·8 per cent. The latter effect was greatest with the alloy containing least aluminium.

Ebert [152] has published a description of apparatus in use at the Physikalisch-Technischen-Reichsanstalt for the measurement of expansion coefficients, and has given results for rhodium and a number of aluminium and magnesium alloys. Up to temperatures of 1100° C. quartz insertions are used to connect the test rod to the measuring device, since its expansion is small and well known. In the work now reported for rhodium, temperatures as high as 1500° C. have been reached. This is well beyond the softening temperature of quartz and so use has been made of sintered corundum. This is one of three refractory

materials of which the expansion coefficients had been measured previously by Ebert and Tingwaldt [153]. The others were magnesia and zirconia. Temperatures as high as 2000° C. were reached, the specimen being heated in a tube of zirconium oxide, 10 to 11 cm. long, of 6·5 to 7 mm. internal diameter and 10·5 mm. outer diameter. This tube had end connexions of a platinum-rhodium alloy, and, after its temperature had been raised sufficiently for the refractory tube to become conducting, it could be taken to 2000° C. by a current of 5 to 6 amperes at 200 volts. The results for the sintered corundum were reproducible up to 1800° C., and this material was considered suitable for use as a comparative standard up to this temperature.

REFERENCES

(140) Hull, A. W., Burger, E. E. and Navias, L. *Phys. Rev.* **49**, 647 (1936).
(141) Robertson, J. M. and Ubbelohde, A. R. *Proc. Roy. Soc.* A, **170**, 241 (1939).
(142) Megaw, H. *Z. Kristallogr.* **100**, 58 (1938).
(143) Owen, E. A. and Yates, E. L. *Phil. Mag.* **21**, 809 (1936).
(144) Marick, L. *Phys. Rev.* **49**, 831 (1936).
(145) Bach, R. *Helv. phys. Acta*, **2**, 95 (1929).
(146) Rosenbohm, E. *Physica*, **5**, 385 (1938).
(147) Ackley, R. A. and Walters F. M., Jr. *Metals & Alloys*, **7**, 314 (1936).
(148) Wells, C., Ackley, R. A. and Mehl, R. F. *Trans. Amer. Soc. for Metals*, **24**, 46 (1936).
(149) Mauer, E. and Schmidt, W. *Mitt. K.-Wilh.-Inst. Eisenforsch.* **2**, 8 (1921).
(150) Hidnert, P. *Bur. Stand. J. Res., Wash.*, **16**, 529 (1936); **17**, 697 (1936).
(151) Hidnert, P. *Bur. Stand. J. Res., Wash.*, **20**, 809 (1938).
(152) Ebert, H. *Phys. Z.* **39**, 6 (1938).
(153) Ebert, H. and Tingwaldt, C. *Phys. Z.* **37**, 471 (1936).

KINETIC THEORY OF THE ELASTICITY OF RUBBER

By H. PELZER

THERE is a great difference between the elasticity of a steel spring and the elasticity of rubber. If we expand a steel spring and then heat the expanded spring its tension will decrease. If we expand a piece of rubber and heat it in the expanded state the tension as a rule will increase.

The reason for this difference is that the elasticity of steel is caused by molecular forces whilst the elasticity of rubber is rather to be compared with the elasticity of a gas compressed by a piston : the elastic forces here are caused by the thermal motion of the molecules.

More mathematically we can express this in terms of the statistical connexion between probability of state and entropy. For a gas a state of large volume is more probable than a state of small volume. If we compress the gas by a piston we produce a state of less probability and the increased force of pressure tries to restore the more probable state.

The elasticity of rubber is quite similar. Rubber consists of long molecular chains, the angle at the carbon atoms not engaged in double bonds being the tetrahedral angle. But as there is free rotation round the single bonds the aspect of most of the chains will not be a stretched one but a rather crumpled one.

In this much more probable crumpled state the ends of the chain will be much nearer to one another than in the very improbable stretched state.

And now it is not difficult to imagine what will happen if one tries to stretch one end of the chain from the other whilst the links of this chain, weighted with carbon atoms, are in thermokinetic motion, the whole chain behaving like a curling snake ; one can almost feel the tension caused by this thermal motion, which will result in the endeavour to bring the ends of the chain as close together as possible. Now suppose we increase the thermal motion by heating. The result will be that the tension resisting our attempt to stretch the chain will increase with temperature, just as rubber was stated at the beginning to behave, and contrary to the behaviour of a steel spring.

To treat this problem mathematically we use the thermodynamic equation expressing the free energy F in terms of the entropy S, temperature T and potential energy U :

$$F=U-TS.$$

The force f connected with an alteration of any coordinate q thus becomes

$$f=-\frac{\partial F}{\partial q}=T\frac{\partial S}{\partial q}-\frac{\partial U}{\partial q}.$$

As $T\,dS=dQ$, the first term denotes the alteration of heat energy Q with alteration of length q. In the case of steel this first term nearly vanishes, whilst, in the case of a rubber sample investigated by Meyer and Ferri [2], the authors have been able to show that all change of energy was due to change of heat energy, so that in this case the second term vanished. (One has to be careful to eliminate heat of crystallization occurring by overstretching.) Of course in general both terms will be different from zero, but the elasticity of steel will be due mainly to the second term $\frac{\partial U}{\partial q}$, whilst the elasticity of rubber is due mainly to the first term $T\frac{\partial S}{\partial q}$.

To express the entropy S by known quantities we make use of the statistical connexion between entropy and probability of state as given by

$$S=\boldsymbol{k}.\log P(q),$$

where $\boldsymbol{k}$ denotes the Boltzmann constant and P is the probability of state expressed in terms of a size coordinate q (e.g. the volume of the gas or the distance of the ends of the chain from one another).

To give an example of the application of these equations, we shall show how to treat by them the simple case of an ideal gas.

The probability of finding a molecule of the gas within a volume V is proportional to this volume :

$$P=c.V;$$

therefore according to our first equation the free energy becomes

$$F=-\boldsymbol{k}T(\log V+\text{const}).$$

As in this case q is represented by the volume V and the force f by the gas pressure p, the equation $f=-dF/dq$ becomes

$$p=-dF/dV=\boldsymbol{k}T/V,$$

i.e. the well-known gas equation.

After this illustration we will now transfer this treatment to the problem of the molecular chain of rubber.

The rubber chain consists of a number N of links. Actually the direction of any one of these links is not quite independent of the direction of the neighbouring links in so far as the neighbouring link must lie on the surface of a cone of angle equal to the angle of valency between adjacent links. But on account of the free rotation there will be practically full independence between the direction of a valence bond and the direction of its third or fourth neighbour. Therefore we can treat the problem in the way adopted in the theory of errors or in the theory of the Brownian movement.

For the moment we simplify the problem by considering the one-dimensional problem instead of the three-dimensional one.

We consider a chain of N links. Any of these links is directed either to the left side or to the right. Therefore the probability that N_1 of these links are directed to the left side and N_2 to the right (where $N_1+N_2=N$) is

$$P_{N_1,N_2}=2^{-N}.N!/N_1!\;N_2!$$

which has a maximum if $N_1=N_2=N/2$. This means that the most probable length of the chain will be zero. Of course this should be so, as there cannot be any preference either for the left side or the right. If we develop the above expression in the neighbourhood of the value $N_1=N_2=N/2$ by use of the Stirling formula we get for large N :

$$P_{\frac{N}{2}+a,\ \frac{N}{2}-a}=P_a=2^{-N}\cdot\frac{N!}{\left(\frac{N}{2}+a\right)!\left(\frac{N}{2}-a\right)!}\simeq \text{const.}\ e^{-2a^2/N}$$

for the probability that there are $(\frac{1}{2}N+a)$ steps to the right and $(\frac{1}{2}N-a)$ steps to the left. Here the length x of the chain is not zero but $2a$ links, i.e. $x=2ab$ if b is the length of one link.*

Substituting $a=x/2b$ we thus get the probability proportional to $e^{-\frac{x^2}{2b^2N}}$ or, generalizing to three coordinates, proportional to

$$\exp\left(-\frac{x^2+y^2+z^2}{2b^2N}\right).$$

Normalizing, we get the probability for a chain having one end at the origin of the coordinate system and the other end at the point $(x, y. z)$ within an element of volume (dx, dy, dz) :

$$P_{x,\,y,\,z}\,dx.dy.dz=\frac{1}{(\sqrt{2N\pi}.b)^3}\cdot e^{-\frac{x^2+y^2+z^2}{2b^2N}}\,dx.dy.dz.$$

Thus in the most probable state the ends of the chain are at zero distance from each other.

Now we have to insert this expression for the probability into the formula for the free energy $F=U-\boldsymbol{k}.T.\log P$, as we did in the case of the ideal gas. Thus we get the tension s of the chain due to entropy forces by differentiation in the x-dimension :

$$s=\boldsymbol{k}.T.\frac{d\log P}{dx}=-\frac{\boldsymbol{k}.T}{b^2N}x,$$

i.e. a quasi-elastic tension opposed to and proportional to the elongation x and proportional to the absolute temperature T.

Reasonable as this result might seem at first sight, nevertheless there is considerable difficulty in interpreting it. For it is impossible to compare our result $\left(\text{tension}=\frac{\boldsymbol{k}T}{b^2N}.x\right)$ with the technical conception of the modulus of

* If we take account of the angle of valency $180-A$ we have to modify the value of b according to a calculation by Eyring, who obtained an expression for the average length of a chain by integration over the free rotations of each of the N valences :

$$\bar{r}^2=L^2[(N+2(N-1)\cos A+2(N-2)\cos^2 A+\ldots+2\cos^{(N-1)}A)]\simeq L^2N\frac{1+\cos A}{1-\cos A}.$$

Putting this equal to $r^2=3Nb^2$ according to the theory of errors we get for the effective length b expressed by the actual length L of the valence bond :

$$b=\frac{L}{\sqrt{3}}\sqrt{\left(\frac{1+\cos A}{1-\cos A}\right)}.$$

elasticity E defined by

$$\text{Tension} = E\frac{D-D_0}{D_0} \quad (D=\text{length},\ D_0=\text{initial length}).$$

The reason for this is clear. According to our formula the tension of a single chain vanishes only if the length x of the chain vanishes, whilst a piece of rubber has a definite length D_0 even if there is no tension on it. Therefore our expression has not yet the dimension of a modulus of elasticity.

According to W. Kuhn[1] the composition of a piece of rubber by molecular chains can be regarded also as a statistical effect only. Although the most probable state of a chain is that of length zero, yet there will of course be many chains of less probable length, i.e. of non-vanishing length r. The mean value of r^2 is not zero but $\overline{r^2}=3b^2N$.

Now if the rubber is expanded—say in the direction z—by the amount $\frac{D-D_0}{D_0}+1$, all the chains will be expanded in the z-direction by this amount whilst the x- and y-components will not alter. So the chains that had previously the length (distance between their ends) zero will not be expanded at all, but chains of a less probable initial length z_0 (i.e. non-vanishing length) in the z-direction will be expanded to the length

$$z=z_0\left(\frac{D-D_0}{D_0}+1\right) \text{ in the } z\text{-direction.}$$

Only these latter chains contribute to the elasticity (as it is their entropy only that is changed by expansion), whereas chains in the most probable state of length (zero) do not contribute to the elasticity at all, as their length will not alter by expansion.

By integration over the entropy of the whole distribution of lengths, Kuhn gets the result for change of entropy by expansion :

$$S_0-S=\frac{3}{2}\boldsymbol{k}.G.\left(\frac{D-D_0}{D_0}\right)^2,$$

where G is the number of chain molecules per unit of volume. Assuming that the volume of rubber does not alter during expansion, so that there is a contraction in the direction of x and y with the expansion in the z-direction, Kuhn gets the modified result

$$S_0-S=\frac{7}{2}\boldsymbol{k}.G.\left(\frac{D-D_0}{D_0}\right).$$

Thus the tension becomes

$$dF/d\left(\frac{D-D_0}{D_0}\right)=7\boldsymbol{k}TG\frac{D-D_0}{D_0}$$

and the modulus of elasticity is

$$E=7\boldsymbol{k}.T.G.$$

If we denote the density by d and the molecular weight of the chain by M we have $G=Ad/M$, where A is the Avogadro number per mol. As the gas

constant R is $R=kA$, the modulus of elasticity according to Kuhn finally becomes

$$E=7RTd/M.$$

As mentioned above, Meyer and Ferri have tested a sample of rubber which behaved " ideally " in so far as the elastic energy was due entirely to change in heat energy. The modulus of elasticity was 5.10^8 dynes/cm^2 for this sample. According to the above formula of Kuhn, the corresponding mean molecular weight should be $M=35000$. Although it looks rather high, this value is quite a possible one. For if two chains get interlocked, the effect might be that of a chain of greater effective length, the two chains counting as one. But on the other hand one might also expect that the effect would be that of shorter effective length of chains, because the lengths of the chains are to be counted only from one connexion to the next, so that the other part of a chain behind the connecting point counts as a different chain. In this respect the theory of Kuhn is not complete, as the connexions of the chains are not taken into account at all, the calculation being carried out on the assumption that the different chains are independent of one another, so that the entropy of the whole piece of rubber can be assumed to be the sum of the entropies of the single chains.

The complete solution of the problem would be nothing less than an adequate treatment of the statistics of networks.

Connecting chains by vulcanizing certainly has the effect of shortening the length of the chain.

Friction of chains (viscosity) involves a time effect, the deformation being partly a plastic irreversible one and partly an elastic reversible one. A rapid expansion will cause a considerable tension but this will decrease with time, as is in fact found in the phenomenon of relaxation.

REFERENCES

(1) Meyer, K. H., Valkó and Susich. *Kolloidzschr.* **59**, 208 (1932).
Guth and Mark. *Mh. Chem.* **65**, 93 (1934).
W. Kuhn. *Kolloidzschr.* **76**, 258 (1936).
(2) Meyer and Ferri. *Helv. Chim. Acta*, **18**, 570 (1935).

SUPERCONDUCTIVITY

By L. C. JACKSON, M.Sc., Ph.D.
H. H. Wills Physical Laboratory, University of Bristol

§ 1. INTRODUCTION

During the period which has elapsed since the appearance of the last article in these *Reports* dealing with superconductivity (1936), the work on this subject has been concentrated rather upon the magnetic and thermal properties of superconductors than upon their electrical properties. This is understandable, as "zero magnetic induction" must now be regarded as the fundamental characteristic of a superconductor and the absence of electrical resistance as a secondary effect. The nature of the "intermediate state", the relationships between magnetic, thermal and electrical properties predicted by the thermodynamical treatment of superconductivity and the behaviour of thin films of superconducting metals have been studied in considerable detail. The present *Report* deals almost entirely with these aspects of the subject.

Some results of a general nature deserve mention. To the list of superconducting elements published in the 1934 *Report* we have now to add lanthanum [1] (4·71° K.), zirconium [2] (0·70° K.) and hafnium [2] (0·35° K.); the figures in brackets give the normal transition temperatures. By employing the adiabatic demagnetization method for the production of very low temperatures, it has been shown that magnesium, copper, gold and bismuth [2] do not become superconducting even when cooled to 0·05° K.

An investigation by Grassmann [3] has still further reduced the upper limit which must be set to the micro-residual resistance of superconductors. By examining the decay of "persistent currents" with a sensitive dynamometer he has shown that the resistance of thin superconducting layers is less than 10^{-14} of the resistance at room temperature for tin and less than 5×10^{-16} for lead. The fact that each progressive refinement in experimental technique still indicates that the electrical resistance of a superconductor is strictly zero is of importance as, by reversing the usual argument, it may be used to prove that the magnetic induction is also exactly zero.

§ 2. THE INTERMEDIATE STATE

Magnetic properties. When a long cylindrical superconductor is placed in a uniform magnetic field with its axis parallel to the direction of the field and the strength of the latter is gradually increased, the "critical" field strength H_c is reached everywhere on the surface of the superconductor at the same instant. The magnetic field penetrates the specimen and the superconductivity is destroyed, the transition $s \rightarrow n$ being quite sharp under these conditions for a pure metal. On the other hand, the presence of a superconductor of any other shape distorts the magnetic field, and on increasing the

strength of the latter the critical field is reached at certain parts of the surface of the specimen earlier than at others. When the field at the surface is everywhere smaller than the critical field, the body is wholly in the superconducting state provided that its temperature is sufficiently low; and similarly when the applied field is everywhere greater than H_c, the body is wholly in the normal state. When, however, the external field at the surface has anywhere reached H_c the superconductivity is destroyed locally and the specimen is in the "intermediate state"

The magnetic properties of the intermediate state have been investigated by Meissner and Ochsenfeld [4] (measurement of the magnetic field around the specimen by a ballistic method), Mendelssohn and Babbitt [5] (*ibid.* using a bismuth wire), de Haas and Guinau [6] (measurement of the induction and magnetizing field inside a sphere by bismuth wires), Rjabinin and Shubnikov [7] (measurement of the induction by a ballistic method), Shoenberg [8] (measurement of the force on a sphere in a slightly inhomogeneous field), Daunt [9] and Shoenberg [10] (measurement of the alternating-field susceptibility of the specimen).

The results for the case of a sphere may be taken as typical. If a sphere is placed in a uniform magnetic field H_1, the internal field H is uniform and equal to $H_1-\frac{4}{3}\pi I$. When the sphere is superconducting $B=0=H+4\pi I$ and the internal field reaches the critical value H_c when $H_1=\frac{2}{3}H_c$. It is actually found that when the applied field reaches $\frac{2}{3}H_c$, the external field at the equator of the sphere becomes equal to H_c and a magnetic induction begins to appear inside the sphere.* As H_1 rises to the value H_c, the induction increases linearly from zero to H_c according to the expression $B=3H_1-2H_c$, the intefnal field H remaining equal to H_c throughout this interval. The sphere is now wholly in the normal state and $B=H_1$ for any further increase of H_1. If the field strength is now reduced to zero, the above path is retraced exactly by B.

On the other hand the sphere may be cooled to a low temperature T_0 in zero magnetic field, and a field too small to destroy the superconductivity at this temperature but greater than $\frac{2}{3}H_c$ at the normal transition point T_0 may be applied and maintained constant. If now the temperature is made to rise, B remains zero inside the sphere until that temperature is reached at which the applied field is equal to $\frac{2}{3}H_c$ (H_c is a function of T). B then increases linearly with further increase of T until it reaches H_c (T), after which it remains constant at this value.

When the external field at the surface of the specimen rises above H_c it may be supposed that the field penetrates the sphere and destroys the superconductivity within a certain volume, determined by the fact that the field must be H_c at the boundary surface between the superconductive and normal regions. A consideration of the equilibrium at the surface of separation of these

* The description refers to the slightly idealized case which the experimental results seem to indicate would occur with an absolutely pure monocrystalline specimen free from strain. The results were obtained with a monocrystalline sphere of tin cooled initially in zero magnetic field to some temperature below the normal transition point. The temperature was subsequently kept constant while a magnetic field was applied and its strength changed as described.

regions inside the specimen shows however that the intermediate state must have a complex structure. It is impossible to find a simple boundary surface for which the field has the value H_c on the surface and is greater than H_c between it and the surface of the sphere. Peierls [11] and F. London [12] have, however, shown that though a *microscopically* complex structure must be assumed, this leads to a *macroscopic* description of the intermediate state which exactly reproduces the observations of de Haas and Guinau. For a detailed study of the intermediate state it is necessary to assume that the specimen breaks up into a series of alternately superconducting and normal laminae parallel to the applied field, the fraction of the total volume occupied by the normal laminae being given by B/H_c.

A thermodynamical treatment of the equilibrium of this proposed laminar structure of the intermediate state has been made by Landau [13] for the case of a plate in a transverse magnetic field. He assumed a " surface tension " between the layers and shows that while the latter are approximately plane and parallel in the interior of the specimen, the normal layers branch repeatedly along the lines of force as they approach the surface. Consequently there will be a very intimate mixture of extremely small superconducting and normal regions immediately below the surface of the specimen. The conditions for a minimum free energy require that the thickness of the laminae near the surface shall be smaller than the depth to which an external magnetic field penetrates into a superconductor (about 10^{-4} cm.; see § 4). For a plate of thickness 1 cm. the greatest thickness of the laminae may be estimated as about 2×10^{-2} cm.

Electrical properties. The microscopically inhomogeneous character of the intermediate state is proved very clearly by the observations of Shubnikov and Nakhutin [14]. They showed that a sphere in the intermediate state has an electrical resistance perpendicular, but not parallel, to the direction of the applied field, in agreement with the presence of alternating superconducting and normal laminae.

The restoration of the electrical resistance of a cylindrical conductor by a transverse magnetic field also furnishes evidence for the existence of an intermediate state. Since with a superconducting circular cylinder the maximum field at the surface is twice as great as the applied field, it is to be expected [15] that the resistance would begin to return when $H_1=\frac{1}{2}H_c$ and then, on the above described view of the structure of the intermediate state, would increase linearly to its full value. The fact that this is even approximately so, clearly shows that superconducting and normal regions must be present and so situated that the current must pass partly through the normal regions.

Actually de Haas, Voogd and Miss Jonker [16] found that the mode of restoration of the resistance of a monocrystalline tin wire in a transverse magnetic field at 2·92° K. is greatly dependent on the strength of the measuring current, the linear rise of resistance only being approached for high current strengths. Furthermore, the first trace of resistance is observed when H_1 is equal to $0{\cdot}58\,H_c$ and not to $\frac{1}{2}H_c$. Later measurements by Misener [17] showed that the field which first restores a resistance is a linear function of ΔT, the

difference between the normal transition temperature and the temperature of the experiment. Extrapolating the straight line gives precisely $0{\cdot}50\,H_c$ for $\Delta T=0$.

Mann, Grayson Smith and Wilhelm [18] have made similar experiments with cylindrical conductors in a wholly superconducting circuit. The first trace of resistance was detected by the decay of the persistent current flowing in the circuit. With a specimen of purest tin the resistance is first restored at $0{\cdot}58\,H_c$ independently of the value of ΔT, but for less pure tin results exactly similar to those of Misener were obtained. The discrepancy between the two series of measurements is the more surprising as Misener's tin wires were prepared from purest tin of the same origin as the Toronto workers' purest specimen (Hilger 10,000). Since the specimens used differed greatly in diameter (Mann and others, $d=3{\cdot}65$ mm.; Misener, $d=0{\cdot}10$ or $0{\cdot}15$ mm.; de Haas and others, $d=0{\cdot}25$ mm.) the explanation may lie in Shoenberg's suggestion [19] that the precise value of H_1/H_c at any temperature may depend on the anomalous properties of the very fine laminae of the intermediate state immediately below the surface of the specimen.

The properties of the intermediate state are also involved in the explanation of the gradual restoration of the resistance of a conductor when the current flowing through it exceeds a certain critical value (Shubnikov and Alexejevski [20]). Since, however, in the case of a cylindrical wire carrying a current the lines of force are circles, the intermediate state cannot have precisely the structure described above. It is beyond the scope of the present article to deal with this problem in detail.

Relaxation phenomena. A puzzling feature of the intermediate state is the various relaxation effects associated with some of the changes of H or T. De Haas, Miss Engelkes and Guinau [21] found that at the beginning of the transition region a considerable time (up to 30 minutes) is required for the induction in a monocrystalline tin sphere to acquire its final value, whether the transition is made by heating in a constant magnetic field or by increasing the field at constant temperature. The time lag decreases with increasing H, becoming zero just before the transition is complete. These relaxation effects cannot be due to any lack of thermal equilibrium as this is established within a few seconds at most. They depend, however, on the shape of the specimen, whether the latter is mono- or polycrystalline and on the manner in which the magnetic field is changed at constant temperature [22].

Keesom and van Laer [23] have shown that the temperature equilibrium in calorimetric experiments is reached with a relaxation time of about 30 seconds when the transition ($s\rightarrow n$) is produced by heating in a constant magnetic field. On the other hand, if the transition is brought about by increasing the applied field no such relaxation effects are observed. Grayson Smith and Mann [24] claim that all these relaxation phenomena can be explained by the decay of persistent currents when a trace of electrical resistance is restored locally. van Laer and Groenewold [25], however, insist that the 30-sec. period associated with thermal effects must be of different origin from the 30-minute period associated with changes of induction. The effects

may, however, be taken to indicate that (1) there is a *weak* coupling between the atomic lattice and the superconducting electrons near the threshold curve, (2) there is a *strong* coupling between the external magnetic field and the superconducting electrons, and (3) there is a *strong* coupling between the normal electrons and the atomic lattice.

§3. THERMAL PROPERTIES

Direct calorimetric measurements. The thermodynamical treatment of superconductivity outlined in the 1936 *Report* has been verified in greater detail and to a greater accuracy than was previously possible so that it is now established that the transition from the superconductive state to the normal state in the presence of a magnetic field is an *ordinary reversible phase transformation* (in the absence of a magnetic field the transformation is of the "second kind"). Keesom and van Laer have carried out an extensive series of calorimetric experiments on tin which provide a much better test of the thermodynamical theory than the earlier work on thallium. The tin specimens used had the form of elongated ellipsoids of revolution with their long axes parallel to the applied magnetic field, a shape more suited for exact comparison with theory than the short cylinder previously used. This shape also restricted the interval of temperature or field strength in which the specimen was in the intermediate state to a small and accurately predictable value.

Using an ellipsoid (axial ratio 5 : 1) for which the field at the equator was $1{\cdot}059\times$ the applied field, Keesom and van Laer [26] measured the latent heat associated with the transition from superconductivity to normal conductivity produced by heating in a constant magnetic field $[(s\rightarrow n)H_0]$. The results, averaged over the small transition interval (0·04°–0·10°), are given in table 1 together with the theoretical values deduced from the formulae:

$$r_1=\int_T^{T_0}\Delta c\,dT-\frac{H^2V}{8\pi},$$

$$r_2=T\int_T^{T_0}\frac{\Delta c}{T}\,dT,$$

(Δc=difference in atomic heats of superconductive and normal states, V=atomic volume, T_0=transition temperature in zero magnetic field), derived from the first and second laws of thermodynamics respectively (Gorter [27]). The agreement between r_1, r_2 and the observed values is evidence of the reversibility of the transition.

Table 1. *Latent heat of superconductivity of tin*

Applied field (gauss)	Observed (cal.)	r_1 (cal.)	r_2 (cal.)
49·5	$43{\cdot}4\times10^{-4}$	$45{\cdot}7\times10^{-4}$	$45{\cdot}6\times10^{-4}$
99·0	70·3	69·5	69·0
138·6	75·3	74·1	75·6

The latent heat associated with the corresponding transition, the destruction of superconductivity by increasing the magnetic field at constant temperature $(s \rightarrow n)\, T_0$, was also investigated for tin [28]. The variation of r with temperature was found to be in excellent agreement with that calculated from

$$r = -\frac{1}{4\pi} \cdot VTH_c \frac{dH_c}{dT}$$

using the data of de Haas and Miss Engelkes [29] for the H_c, T curve of tin.

In addition, van Laer and Keesom [30] have been able to prove experimentally that the latent heats associated with the transitions $(s \rightarrow n)\, T_0$ and $(n \rightarrow s)\, T_0$ are exactly equal, thus completing the evidence for the reversibility of the transition from superconductivity to the normal state in the presence of a magnetic field.

The atomic heat of tin has been determined by Keesom and van Laer [31] from 3·94° K. to 1·14° K. (*a*) in the absence of a magnetic field and (*b*) in the presence of fields of 138·6 and 299·0 gauss. In case (*b*) the specimen was in the normal state throughout and the observations lie on a smooth curve which is a continuation of the atomic-heat curve at higher temperatures. In case (*a*) there is a discontinuous increase of the atomic heat on passing the transition temperature. The differences in the atomic heats in the superconducting and normal states, $\Delta C = C_s - C_n$, were found to agree well with the thermodynamic formula

$$\Delta C = \frac{VT}{4\pi}\left[\left(\frac{dH_c}{dT}\right)^2 + H_c \frac{d^2H_c}{dT^2}\right]$$

The atomic heat in the superconducting state follows approximately the law $C_s = 464{\cdot}5\,(T/\Theta)^3$ with $\Theta = 140$. On the other hand it is possible to represent the values of C_n by the formula

$$C_n = 464{\cdot}5\,(T/\Theta')^3 + \gamma T$$

with $\Theta' = 185$ and $\gamma = 0{\cdot}00040$. The first term represents the contribution of the lattice vibrations to the atomic heat and the second term that of the conduction electrons. Using Sommerfeld's expression, this value of γ corresponds to 7·2 " free electrons " per atom.

Indirect determinations from magnetic data. These direct calorimetric measurements have received valuable confirmation from the work of Daunt, Horseman and Mendelssohn [32]. The equilibrium curves * of thallium, indium, tin and lead were redetermined and the entropy differences Δs between the superconducting and normal states were calculated from the formula

$$\Delta s = \frac{VH_c}{4\pi} \cdot \frac{dH_c}{dT}.$$

The values of ΔC $(= T\, d\Delta s/dT)$ obtained for tin are in excellent agreement with Keesom and van Laer's observations.

* The term " equilibrium curve " is used by these authors to denote the H_c, T curve, the boundary curve separating the superconducting and normal regions as deduced from induction measurements, the term " threshold field curve " being restricted to that deduced from resistance measurements.

For a substance which follows the formulae given above for C_s and C_n, the slope of the $\{\Delta s, T\}$ curve at the absolute zero is equal to $-\gamma$. The values obtained by Daunt, Horseman and Mendelssohn in this way are given in table 2.

Table 2

Metal	γ
Thallium	$2{\cdot}8\times10^{-4}$ cal./mol.
Indium	3·5 ,,
Tin	3·95 ,,
Lead	7·0 ,,

For these substances the contribution of the superconducting electrons to the atomic heat in the absence of a magnetic field can be calculated from

$$C_{s.\,el.}=\Delta C+\gamma T.$$

$C_{s.\,el.}$ is shown to remain very small until the temperature rises to 1 to 3° K. and then to rise rapidly to the normal transition temperature. This contribution to the atomic heat may be regarded as the heat required to lift electrons from the superconductive to the normal state, so producing a gradual change in the proportion of superconducting to normal electrons from 100 per cent at 0° K. to 0 at the transition point.

Misener [33] has recently made a further contribution to this section of the subject by showing that the equilibrium curves can be represented accurately by the first few members of a series of ascending powers of the temperature. The various thermodynamical quantities involving dH_c/dT and d^2H_c/dT^2 can then be obtained by direct calculation with a greater accuracy than was possible with the graphical methods previously used. The results so obtained do not differ greatly from those of Daunt, Horseman and Mendelssohn but the curves, particularly those of $C_{s.\,el.}$, are smoother.

§ 4. SUPERCONDUCTIVITY OF THIN FILMS

The fact that the magnetic induction is zero in the interior of any massive pure superconductor implies that all electric currents flowing must be superficial. In reality such currents must penetrate to a certain finite depth which must, however, be very small compared with 1 cm. Quantitative data on the manner in which a current or an external magnetic field * penetrates into the surface layer of a superconductor are of great importance as they may possibly provide the clue to the still unsolved problem of the origin of the zero permeability of superconductors. The fact that the gyromagnetic ratio of superconductors is equal to $e/2m.c$ [34] indicates that the zero permeability is to be associated with electron *currents* and not with electron *spin*. Our other information on this problem is very scanty but it is probable that experiments

* It has been shown by H. London [42] that an electric field does not penetrate even into the superficial layer.

on superconductors whose dimensions are comparable with the expected depth of penetration will prove valuable.

In the early work on the superconductivity of thin films—Sizoo and Kamerlingh Onnes [35] (sputtered films), Misener and others [36] (lead and tin electrolytically deposited on non-superconducting metallic supports)—it was practically impossible to separate the true size-effect from that due to the structure of the films or the strain present. Shalnikov's work [37] on films of tin and lead prepared by evaporation on to a plane glass surface at 4·2° K. in a high vacuum is more satisfactory but still not completely reproducible. These measurements show that (1) superconductivity can occur in films only 15 atoms thick, (2) the transition temperature in the absence of a magnetic field is the same for annealed films as for the bulk metal, but may differ considerably from this value with unannealed films, (3) the critical current strength at which the normal resistance is restored is quite sharp, but lower than that expected on Silsbee's hypothesis, and (4) the longitudinal threshold magnetic field is considerably greater than that of the bulk metal.

It has been shown by Appleyard and Bristow [38] that mercury films produced by the Lovell-Appleyard technique (evaporation on to a pyrex glass surface cooled to 64° K. under conditions of extreme cleanliness and high vacuum with subsequent annealing at 90° K.) are coherent and uniform if more than 4×10^{-6} cm. thick, and, moreover, possess a resistivity which closely approaches that of the bulk metal. The superconducting properties of these films have now been investigated [39], a very careful study having been made of the variation of the critical longitudinal field with temperature and film thickness.

If H_f is the critical field for the film and H_c that for the bulk metal, the ratio H_f/H_c is always greater than unity and increases rapidly for any particular film thickness as the temperature rises to the normal transition point. Since it is to be expected on any theory that H_f will increase rapidly when the thickness of the film is so reduced as to be comparable with the depth of field penetration, these results indicate that this depth varies with the temperature, becoming very large near the transition point and possibly infinite there. The experimental results alone do not furnish a definite value for the penetration depth, but by combining all the results for various temperatures and for films differing in thickness by a factor of 50, it is possible to draw a curve of the relative depth of penetration $\lambda/\lambda_{2\cdot5}$ ($\lambda_{2\cdot5}$=depth of penetration at the arbitrarily chosen standard temperature 2·5° K.) as a function of the temperature. The excellent consistency of the results derived from the different films shows that one is dealing with a true size-effect.

If an exponential penetration is assumed it is possible to deduce the relationship

$$H_f/H_c=1+\lambda/d \quad (d=\text{thickness of film})$$

for very thick films. The observations then lead to a value of $\lambda=1\cdot2\times10^{-4}$ cm. for mercury at 2·5° K. A similar expression had been found by Pontius [40] to fit his measurements on very fine lead wires, and his calculated value of λ is of the same order of magnitude.

The above value of λ may be compared with that calculated from the formula of Londons' theory, $\lambda = mc^2/4\rho ne^2$, which gives 2×10^{-6} cm. if the number of "free electrons" in the metal (n) is put equal to the number of valency electrons. If one reverses the procedure and calculates n from the value of λ deduced from the observations on very thick films one finds that n becomes zero at T_0 and never exceeds 0·05 even at very low temperatures. The reason for the discrepancy between this conclusion and that drawn by Daunt, Horseman and Mendelssohn (§ 3) has yet to be found.

A further estimate of λ has been made by Shoenberg [41] from measurements of the magnetic properties of superconducting mercury colloids. For a mercury-in-gelatine colloid (average particle diameter 10^{-4} cm.) the ratio χ/χ_0 (χ=magnetic moment per unit field per unit volume of colloid, χ_0=the same for a large sphere $= -\frac{3}{8\pi}$) was always less than unity and fell rapidly towards zero at the transition point. For very small values of r (=radius of particle) it can be shown that the London theory leads to

$$\chi/\chi_0 = r^2/15\lambda^2.$$

The values of $\lambda/\lambda_{2\cdot5}$ at various temperatures deduced from this expression lie very well on the curve published by Appleyard and others for thin films of mercury as described above.

REFERENCES

Collected accounts of superconductivity published since 1936 :—

Steiner and Grassmann. *Supraleitung.* (Vieweg, 1937.)

Shoenberg. *Superconductivity.* (Cambridge University Press, 1938.) An excellent general account for the English reader.

London, F. *Une conception nouvelle de la supraconductibilité.* (Hermann et Cie 1937.) A review of the phenomena of superconductivity from the point of view of the phenomenological theory of F. and H. London.

(1) Mendelssohn and Daunt. *Nature, Lond.*, **139**, 473 (1937).
Shoenberg. *Proc. Camb. Phil. Soc.* **33**, 577 (1937).

(2) Kürti and Simon. *Proc. Roy. Soc.* A, **151**, 610 (1935).

(3) Grassmann. *Phys. Z.* **37**, 569 (1936).

(4) Meissner and Ochsenfeld. *Naturwissenschaften*, **21**, 787 (1933). (See also Meissner and Heidenreich, *Phys. Z.* **37**, 449 (1936): connected account of the Meissner Effect.)

(5) Mendelssohn and Babbitt. *Proc. Roy. Soc.* A, **151**, 316 (1935).

(6) de Haas and Guinau. *Physica*, **3**, 182, 534 (1936).
de Haas and Miss Engelkes. *Ibid.* **4**, 325 (1937).
de Haas, Miss Engelkes and Guinau. *Ibid.* **4**, 595 (1937).

(7) Rjabinin and Shubnikov. *Phys. Z. Sowjet*, **5**, 641 (1934); **6**, 557 (1934).

(8) Shoenberg. *Proc. Roy. Soc.* A, **152**, 10 (1935); **155**, 712 (1936).

(9) Daunt. *Phil. Mag.* **24**, 361 (1937).

(10) Shoenberg. *Proc. Camb. Phil. Soc.* **33**, 559, 577 (1937).

(11) Peierls. *Proc. Roy. Soc.* A, **155**, 613 (1936).

(12) London, F. *Physica*, **3**, 450 (1936).

(13) Landau. *Nature, Lond.*, **141**, 688 (1938); *Phys. Z. Sowjet*, **11**, 129 (1937). (See also Shoenberg, *Superconductivity*, 33–35.)

(14) Shubnikov and Nakhutin. *Nature, Lond.*, **139**, 589 (1937).

(15) von Laue. *Phys. Z.* **38**, 793 (1932).

(16) de Haas, Voogd and Miss Jonker. *Physica*, **1**, 281 (1933).
Misener. *Proc. Camb. Phil. Soc.* **35**, 95 (1939).

(17) Misener. *Proc. Roy. Soc.* A, **168**, 43 (1938).
(18) Mann, Grayson Smith and Wilhelm. *Phys. Rev.* **54**, 758 (1938).
(19) Shoenberg. *Superconductivity*, 38.
(20) Shubnikov and Alexejeski. *Nature, Lond.*, **138**, 804 (1936).
(21) de Haas, Miss Engelkes and Guinau. *Physica*, **4**, 595 (1937).
(22) Mendelssohn and Pontius. *Physica*, **3**, 327 (1936); *Nature, Lond.*, **138**, 29 (1936).
Shubnikov, Chotkewitsch, Schepelew and Rjabinin. *Phys. Z. Sowjet*, Sondernummer, June 1936; **10**, 165 (1936).
(23) Keesom and van Laer. *Physica*, **3**, 173 (1936); **4**, 499 (1937); **5**, 986 (1938).
(24) Grayson Smith and Mann. *Phys. Rev.* **54**, 766 (1938).
(25) van Laer and Groenewold. *Physica*, **6**, 473 (1939).
(26) Keesom and van Laer. *Physica*, **3**, 371 (1936).
(27) Gorter. *Arch. Mus. Teyler.* **7**, 378 (1933).
Gorter and Casimir. *Physica*, **1**, 306 (1933).
(28) Keesom and van Laer. *Physica*, **4**, 487 (1937).
(29) de Haas and Miss Engelkes. *Physica*, **4**, 325 (1937).
(30) Keesom and van Laer. *Physica*, **5**, 993 (1938).
(31) Keesom and van Laer. *Physica*, **5**, 193 (1938).
(32) Daunt, Horseman and Mendelssohn. *Phil. Mag.* **27**, 754 (1939).
(33) Misener. Private communication.
(34) Kikoin and Goobar. *C.R. Acad. Sci., U.R.S.S.*, **19**, 249 (1938).
(35) Kamerlingh Onnes and Sizoo. *Proc. Acad. Sci., Amst.*, **28**, 971 (1925).
(36) Burton, Wilhelm and Misener. *Trans. Roy. Soc. Can.* **28**, III, 65 (1934).
Misener and Wilhelm. *Trans. Roy. Soc. Can.* **29**, 1 (1935).
Misener, Grayson Smith and Wilhelm. *Trans. Roy. Soc. Can.* **29**, 13 (1935).
Misener. *Canad. J. Res.* **14**, 25 (1936).
(37) Shalnikov. *Nature, Lond.*, **142**, 74 (1938).
(38) Appleyard and Bristow. *Proc. Roy. Soc.* A, **172**, 530 (1939).
(39) Appleyard, Bristow, London and Misener. *Proc. Roy. Soc.* A, **172**, 540 (1939).
(40) Pontius. *Phil. Mag.* **24**, 787 (1937).
(41) Shoenberg. *Nature, Lond.*, **143**, 434 (1939).
(42) London, H. *Proc. Roy. Soc.* A, **155**, 102 (1936).

X RAYS AND CRYSTALS

Edited by J. MONTEATH ROBERTSON, M.A., PH.D., D.Sc.,
The University, Sheffield

I. Organic structures. By J. M. Robertson
II. The structures of silicates. By W. H. Taylor
III. The examination of ternary alloy systems by x rays. By H. Lipson
IV. Order-disorder phenomena in alloys. By E. T. Goodwin

IN a previous *Report* [22] the methods of x-ray crystal analysis were discussed in some detail, with special reference to complex molecular structures. The present series of articles has been designed, on the other hand, to give a survey of the results of crystal analysis in a few selected fields. No complete account of all recent advances has been attempted. The general field, covering as it does nearly the whole subject matter of chemistry and including as well many substances so complex and intractable as to make purely chemical investigation difficult, is much too wide for a single report. Instead, a few topics have been selected which, either on account of their intrinsic interest or the importance of their applications, have attracted much attention in recent years.

We wish to express our thanks to Professor W. L. Bragg for assistance and advice, and to the Oxford University Press, the Royal Society, and the Chemical Society for permission to reproduce certain diagrams.

I. ORGANIC STRUCTURES

BY J. MONTEATH ROBERTSON

Introduction. The earlier x-ray work on organic crystals verified in a general way the chemists' picture of the space arrangement of the atoms in carbon compounds. A few supposed anomalies, such as a pyramidal disposition of valency bonds about the central carbon atom in pentaerythritol, have long ago been cleared up in favour of the chemists' ideas. Indeed, physical verification of the fundamental stereochemical conceptions was hardly necessary, because the vast structure of organic chemistry could scarcely have been built on faulty foundations. But the relative positions of the atoms in a molecule constitute only one aspect of its structure, and the importance of fully quantitative x-ray investigations of organic compounds is not diminished but rather increased because of the existing detailed knowledge of their properties. The new information, which forms the experimental basis of modern structural chemistry, includes quantitative measurement of bond lengths and valency angles, and the manner of intermolecular approach and grouping in the solid state. We

consider these two aspects briefly in turn. At least two other important branches in the organic field have recently been extensively studied, but must be deferred for a later Report—the stereochemistry of complex organic metallic compounds, and the structures of proteins and complex biological materials.

Bond lengths in organic molecules. The length of the carbon-carbon single bond is always observed to be within 0·02 or 0·03 A. of the diamond value of 1·54 A. Fully quantitative X-ray investigations confirming this value in saturated aliphatic compounds are not very numerous, but they have been supplemented recently by a large number of electron-diffraction studies [11] which show that the approximate constancy of the single bond is maintained not only in the paraffins but also in cyclic aliphatic compounds, including even the three-membered ring of cyclo-propane, where the bond angle is distorted to 60°. It is true that a number of X-ray investigations have been reported in recent years (pentaerythritol [8], pentaerythritol tetra-acetate [5], succinic acid [28]) which indicate a C—C bond value around 1·52 A. rather than 1·54 A., but the deviations in general lie within the possible experimental error.

In a similar way the length of the carbon-carbon double bond has been determined as 1·33–1·34 A. by spectroscopic and electron-diffraction studies in ethylene [11], while accurate spectroscopic studies of acetylene [6] give the length of the carbon-carbon triple bond as 1·20 A.

A similar constancy is shown by the normal covalent-bond distances between pairs of other atoms such as nitrogen, oxygen, fluorine, silicon, phosphorus, etc., and when such bonds unite two different atoms, such as carbon and nitrogen, the bond distance is in general the mean of the individual C—C and N—N distances. This additivity has made it possible to draw up a table of "covalent radii" for atoms, and the figures are given in recent works [11], [12]. The following are examples for the first-row elements:

	C	N	O	F
Single bond ..	0·77 A.	0·70 A.	0·66 A.	0·64 A.
Double bond ..	0·67	0·61	0·57	0·55
Triple bond ..	0·60	0·55	0·51	—

These radii apply to structures like ethane, ethylene, and acetylene, which can be formulated definitely and satisfactorily in terms of a single valence-bond structure. But for many organic compounds, perhaps the majority, no such simple formulation is possible. Two or more structures must be considered, and the actual properties of the substance may not be represented completely by either. The best examples occur among the aromatic compounds, as in the Kekulé structures for benzene. No single representation in terms of single and double bonds is satisfactory. Many chemical theories of such intermediate structures have been developed, and more recently the quantum-mechanical idea of resonance has been applied to these complex molecules by Slater [27], Hückel [7], and Pauling [10].

The most definite experimental approach to this problem lies in the accurate measurement of bond lengths and valency angles for the molecule concerned, from which data the contributions of the several structures to the normal state

of the molecule may often be deduced. Thus in benzene the C—C distance is 1·39 A., which may be considered to represent 50 per cent double-bond character. The distance is much nearer to the double-bond value of 1·33 A. than to the single-bond value of 1·54 A. because of the stabilization due to the resonance energy. In graphite [2] (C—C=1·42 A.) naphthalene and anthracene [17] (C—C=1·41 A.) and in oxalic acid [19] (C—C=1·43 A.) the measurements indicate a smaller amount of double-bond character. The relation between bond character and distance has been discussed by Pauling and Brockway [13], Penney [15], and Robertson [21].

The dibenzyl-azobenzene series. The X-ray analyses of this series, recently completed [18], [25], show the effect of conjugation between the central multiple bond and the benzene rings. The structures are extremely complex and the

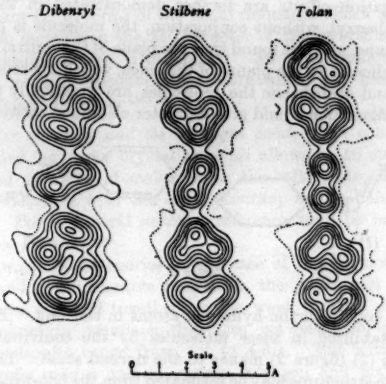


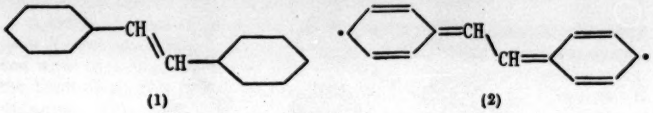
Figure 1.

distance measurements are subject to an uncertainty of about ±0·03 A. Some of the electron-density contour maps, from which the measurements are deduced, are illustrated in figure 1. Not all of the atoms are resolved in these two-dimensional projections, but the form of the six-membered benzene rings (highly inclined to the projection plane) and the pair of connecting atoms can be clearly recognized.

In all cases the benzene rings prove to be approximately regular plane hexagons of side 1·39 A., and the distances between the central pairs of atoms appear to have approximately normal values: C—C=1·54 A. in dibenzyl, C=C=1·33 A. in stilbene, C≡C=1·19 A. in tolan, N=N=1·23 A. in azobenzene. The measurements are not very accurate, but it can be said that the effect of resonance on the length of the multiple bond is slight. But in the "single" bonds, joining the central pairs of atoms to the rings, large resonance effects

are observed, except in dibenzyl, where the $C_{aromatic}—C_{aliphatic}$ distance of 1·49 A., although somewhat less than the normal C—C value of 1·54 A., is probably not a real contraction, and may be largely accounted for by experimental errors. In stilbene, however, the $=C—C_{ar}$ distance of 1·44 A. is definitely below the single-bond value, while in tolan the bond has contracted further to 1·40 A. In azobenzene the N—C link is about 1·41 A., again below the normal N—C value of 1·47 A. It may be noted that in azobenzene the intermediate link of 1·41 A. is somewhat nearer to the N—C single-bond value of 1·47 A. than to the N—C double-bond value of 1·28 A., while in the analogous stilbene structure, the intermediate link of 1·44 A. is midway between the single bond (1·54) and the double bond (1·33) values. This probably indicates that in azobenzene there is a greater tendency for the multiple bond to remain between the nitrogen atoms.

These conjugation effects are further demonstrated in the shape of the molecule. In dibenzyl, without conjugation, the molecule is presumably free to assume any shape, and it is found that the plane of the central $—CH_2—CH_2—$ zig-zag is perpendicular to the planes of the rings, which are separated in space. But in stilbene and azobenzene the molecules are practically plane, although the dibenzyl configuration would give a greater clearance between the ethylene



(1) (2)

Figure 2.

hydrogen atoms and the ortho-hydrogen atoms in the ring. The planar form is presumably stabilized in these molecules by the contribution which the excited structure (2) (figure 2) makes to the normal state. The relative contributions of these structures may be estimated from the bond-distance measurements mentioned above.

X-ray measurements on the recently isolated cis-form of azobenzene [23] indicate that the contribution of structure (2) is decreased in that compound because steric requirements prevent the rings being coplanar. Consequently the C—N link is lengthened to about 1·45 A.

Hexamethylbenzene. A recent re-determination of this structure [4] by the most accurate methods has cleared up one important point, the length of the methyl-group bond to aromatic carbon atoms. As no probable alternative structure can be written for the CH_3 group, this bond length should have the normal 1·54 value. But lower figures, from 1·47 to 1·49, had been reported for durene [16], hexamethylbenzene [9], and dibenzyl [18], mentioned above. With the volatile crystals of durene and dibenzyl it is extremely difficult to make an accurate determination, but hexamethylbenzene is more stable and is also a more favourable structure for accurate analysis, all the atoms lying

nearly in one crystallographic plane. The new determination has led to a methyl-group bond distance of 1·53±0·02 A., in agreement with theory. Whether this figure also applies in a less fully substituted benzene ring cannot be stated with complete certainty, as it has not yet been possible to revise the durene or dibenzyl figures.

The methylacetylene and tolan anomalies. Although the methyl-group bond adjacent to a benzene ring or an ethylenic linkage appears to have approximately the normal single-bond length of 1·54 A., yet there is a definite contraction when it adjoins a triple C≡C link. This has been established in methylacetylene [1], dimethylactylene and dimethyldiacetylene [14], where the contraction is about 0·08 A. In tolan [25] also the "single" bond between the triple bond and the aromatic nucleus shows an additional contraction from the stilbene resonance value of 1·45 A. down to 1·40 A. These results are difficult to account for in terms of the usual resonance effect, and are probably due rather to an increase in the 2*s* character of the hybrid *s*–*h* bond orbital, after the formation of the triple bond. An additional effect, involving resonance with certain unconventional structures, has been examined by Pauling, Springall and Palmer [14].

Intermolecular grouping and the hydrogen bond. The complete picture of structure afforded by X-ray crystal analysis allows us to study the method of approach and grouping of molecules in the solid state, and to determine intermolecular distances with the same accuracy as the bond lengths within the molecules. For a general account reference may be made to a recent article in *Science Progress* [24].

There are two main categories in the case of organic compounds. With hydrocarbons and other inactive molecules, the packing is in general loose and the molecules are held together by the weak residual or van der Waals forces. The minimum approach distances between the carbon atoms of neighbouring molecules range from 3·4 to 4·0 A. But with molecules containing hydroxylic groups and in certain nitrogen and fluorine compounds a much closer minimum approach of from 2·5 to 3 A. is observed. Such molecules are "associated" and remain together in pairs or in small groups in the liquid and sometimes in the vapour state.

The "bonds" which hold these molecules together have one feature in common. They all involve a hydrogen atom, and the nature of these hydrogen bonds, or hydrogen bridges as they are called, has been the subject of much recent research. Here the X-ray diffraction method is at an initial disadvantage because of its inability to locate the hydrogen atom, but accurate measurements can be made of the positions of the oxygen and other atoms. Such determinations have recently been made for resorcinol [20], oxalic acid [19], pentaerythritol [8] and other structures [3].

With regard to the mechanism of the bond, information has recently been obtained by studying the effect of replacing the hydrogen by deuterium [26]. When this is done in hydrocarbons and other crystals not involving the intermolecular hydrogen bond, the general effect is a contraction in volume on

substituting deuterium, which is in keeping with the smaller zero-point energy and consequent smaller space requirement of the deuterium atom. But with hydroxylic compounds involving a hydrogen bond, the effect is reversed, and the crystal spacings concerned show a small expansion on substituting deuterium. In general the expansion effect appears to be greater for the shorter (i.e. stronger) hydrogen bonds, as in oxalic acid dihydrate. Intermediate expansions are observed for $NaHCO_3$, succinic and benzoic acids, and α-resorcinol. For oxalic acid dihydrate the effect is directional, the expansion occurring along the deuterium bonds.

To explain this isotope effect it appears to be necessary to assume a certain amount of intermolecular resonance, at any rate for the shorter types of hydrogen bond. The hydrogen atom, with its greater zero-point energy, will facilitate penetration of the energy barrier between the alternative molecular structures, and should consequently lead to stronger bonds and shorter distances than in the case of deuterium. The expansion observed for deuterium compounds would thus be explained. But this theory is provisional, and further work on the problem is being undertaken.

REFERENCES

(1) Badger and Bauer. *J. Chem. Phys.* **5**, 599 (1937).
(2) Bernal. *Proc. Roy. Soc.* A, **106**, 749 (1924).
(3) Bernal and Megaw. *Proc. Roy. Soc.* A, **151**, 384 (1935).
(4) Brockway and Robertson. *J. Chem. Soc.* 1324 (1939).
(5) Goodwin and Hardy. *Proc. Roy. Soc.* A, **164**, 369 (1938).
(6) Herzberg, Patat and Spinks. *Z. Phys.* **92**, 87 (1934).
(7) Hückel. *Z. Phys.* **70**, 204 (1931); **83**, 623 (1933).
(8) Llewellyn, Cox and Goodwin. *J. Chem. Soc.* 883 (1937).
(9) Lonsdale. *Proc. Roy. Soc.* A, **123**, 494 (1929).
(10) Pauling. *J. Amer. Chem. Soc.* **53**, 1367, 3225 (1931); **54**, 988, 3570 (1932).
(11) Pauling and Brockway. *J. Amer. Chem. Soc.* **59**. 1223 (1937).
(12) Pauling and Brockway. *Proc. Nat. Acad. Sci., Wash.*, **18**, 293 (1932).
(13) Pauling, Brockway and Beach. *J. Amer. Chem. Soc.* **57**, 2705 (1935).
(14) Pauling, Springall and Palmer. *J. Amer. Chem. Soc.* **61**, 927 (1939).
(15) Penney. *Proc. Roy. Soc.* A, **158**, 306 (1937).
(16) Robertson. *Proc.. Roy Soc.* A, **142**, 659 (1933).
(17) Robertson. *Proc. Roy. Soc.* A, **140**, 79 (1933); **142**, 674 (1933).
(18) Robertson. *Proc. Roy. Soc.* A, **146**, 473 (1934); **150**, 348 (1935).
(19) Robertson. *J. Chem. Soc.* 1817 (1936).
(20) Robertson. *Proc. Roy. Soc.* A, **157**, 79 (1936); **167**, 122 (1938).
(21) Robertson. *J. Chem. Soc.* 131 (1938).
(22) Robertson. *Rep. Progr. Phys.* **4**, 332 (1938).
(23) Robertson. *J. Chem. Soc.* 232 (1939).
(24) Robertson. *Sci. Progr. Twent. Cent.* **33**, 663 (1939).
(25) Robertson and Woodward. *Proc. Roy. Soc.* A, **162**, 568 (1937); **164**, 436 (1938); **171**, 398 (1939).
(26) Robertson and Ubbelohde. *Proc. Roy. Soc.* A, **170**, 222, 241 (1939).
(27) Slater. *Phys. Rev.* **37**, 489 (1931).
(28) Verweel and MacGillavry. *Nature, Lond.*, **142**, 161 (1938).

II. THE STRUCTURES OF SILICATES

By W. H. TAYLOR.
Physics Department, College of Technology, Manchester

§1. INTRODUCTION

It is obviously impossible in a Report of this kind to give a detailed description of any of the large number of silicate structures which have been fully analyzed by X-ray methods. Fortunately an authoritative and fully illustrated account is available in W. L. Bragg's *Atomic Structure of Minerals* (Cornell University Press and Oxford University Press, 1936), which includes structure determinations published up to the end of 1936. References to this book will appear in the text in the form "Minerals", with a reference to the chapter.

Possibly the most striking single achievement of X-ray crystallography has been the reduction to order of the vast mass of data previously accumulated in chemical and mineralogical investigations of the silicates. The new technique has demonstrated the soundness of the mineralogist's classifications on the basis of physical properties, and has replaced the chemist's attempt to base silicate chemistry on hypothetical silicic acids by an entirely new conception founded on the geometrical properties of the collections of atoms and ions in the crystal.

The fundamental unit of structure in all silicates is the SiO_4 group, which always has the same dimensions (within small limits), those of a regular tetrahedron of edge 2·6 A. formed by stacking the four oxygen atoms with the silicon at the centre. The forces which bind together the atoms of the SiO_4 group are stronger than any others in the crystal, and are partly covalent; the group as a whole may, however, be considered as an anionic unit $(SiO_4)^{4-}$. The SiO_4 groups may exist as separate and distinct units in the structure, or they may share corners (i.e. oxygen atoms) in various ways ("Minerals", chapter IX). The classification of silicates now generally adopted is based on the type of silicon-oxygen linkage on which the structure is built.

The other constituents such as additional oxygen atoms, hydroxyl groups, water molecules, and cations (Na^{+1}, Ca^{+2}, Mg^{+2}, etc.) are arranged with the SiO_4 groups in such a way as to produce a mechanically stable and electrically neutral structure, and in order to understand the complexities of silicate chemistry the two following facts must be kept in mind. *First*, the aluminium atom plays a very special rôle, for it may appear in either tetrahedral or octahedral coordination with oxygen atoms.* The AlO_4 group has the same dimensions as the SiO_4 group (at least to a first approximation) and may replace

* Reference will be made in §2 to the structures in which the aluminium atom has five-fold coordination.

it in the various structure-types, the necessary adjustment of electrical charge being made elsewhere in the structure when Al^{3+} replaces Si^{4+}. When in octahedral coordination the aluminium atom behaves as a normal trivalent cation. *Secondly*, the possibility of isomorphous replacement, which is of frequent occurrence, is governed almost entirely by similarity in size of the atoms concerned; only to a slight extent by similarity in electric charge; and hardly at all by chemical considerations.

These basic principles of silicate structure will be illustrated in the sections which follow.

§2. STRUCTURES CONTAINING SEPARATE TETRAHEDRA OR COMPLEX SILICON-OXYGEN GROUPS

The structures of a large number of important minerals contain separate SiO_4 groups, and many others are based on complex but "self-contained" groups formed by the sharing of tetrahedron corners by two, three, four or six SiO_4 groups, as illustrated in figure 1. Many of these were amongst the

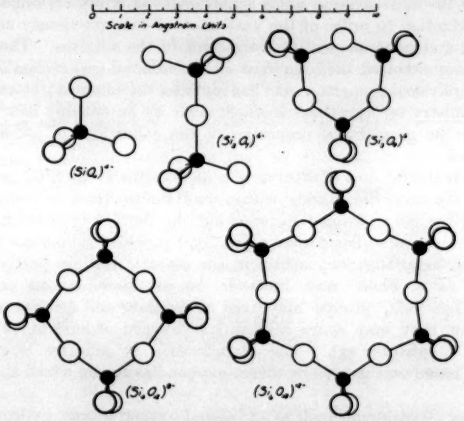


Figure 1. Forms of separate silicon-oxygen groups.

(From *Atomic Structure of Minerals*, by W. L. Bragg (Cornell University Press and Oxford University Press).)

first silicates to be analyzed. For details, see "Minerals", chapters X, XI, XII.

Recent work on this type of structure has been concerned chiefly with the relationships between silicates and other structures built up from tetrahedral anionic groups of oxygen atoms firmly bound to small highly charged

cations such as P^{5+}, As^{5+}, Ge^{4+}. In 1931 Goldschmidt [10] described germanates corresponding structurally to many of the silicate types, and recent detailed analyses of arsenates and phosphates have brought to light very striking similarities with the silicates. Thus, for example, the structure of titanite [77] $CaTiO(SiO_4)$ appears in [66] tilasite $CaMgF(AsO_4)$ and in [36] durangite $NaAlF(AsO_4)$; skorodite $Fe(H_2O)_2(AsO_4)$ probably [68] has the structure [69] of norbergite $Mg_3(OH, F)_2(SiO_4)$; the structure of andalusite $Al(AlO)SiO_4$ with separate SiO_4 groups and five-fold coordination of one-half of the aluminium atoms [70], is exactly reproduced in libethenite [64] $Cu(CuOH)PO_4$, olivenite [27] $Cu(CuOH)AsO_4$, and adamine [35] $Zn(ZnOH)AsO_4$; in the cubic pyrophosphates MP_2O_7(M=Si, Ti, Zr, Sn, Hf, or U) the group P_2O_7 is found [39]; magnesium pyrophosphate [5] is similar to thortveitite [76] $Sc_2Si_2O_7$; and in $Al(PO_3)_3$ there is a ring of four PO_4 groups [59].

The full significance of these relationships appears when a completely systematic classification of mineral structures is attempted, as has been done by Machatschki [40], Náray-Szabó [51], Niggli [55], Laves [38], and Strunz [67]; limitation of space prohibits any account of this work.

§3. CHAIN STRUCTURES. PYROXENES, AMPHIBOLES, AND ASBESTOS

The structures of the pyroxenes and amphiboles are based on the endless single and double chains (respectively) of linked SiO_4 groups illustrated in

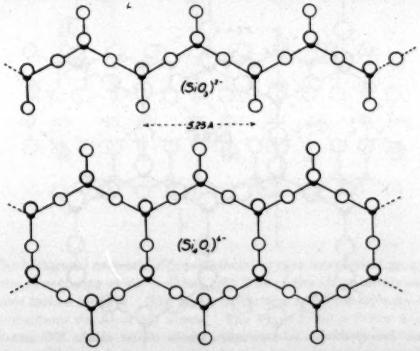


Figure 2. Forms of silicon-oxygen chains. The upper single chain is found in pyroxenes, the lower double chain in amphiboles.

(From *Atomic Structure of Minerals*, by W. L. Bragg (Cornell University Press and Oxford University Press).)

figure 2. The fibrous growth of minerals of these groups, and their characteristic cleavages, are direct consequences of the arrangement of these strong

endless chains with lengths all parallel to the crystallographic *c*-axis. For details of the fascinating chemical and mineralogical relations of these crystals, see " Minerals ", chapter XIII.

Relatively few recent investigations have been concerned with structures of this type, in spite of their obvious importance in relation to the uses of " asbestos ", which may be either amphibole (ideal formula, e.g. $(OH)_2Ca_2Mg_5(Si_4O_{11})_2$) or chrysotile $(OH)_8Mg_6Si_4O_{11}H_2O$, often found in association with the " platy " mineral antigorite in serpentine rocks. Selfridge [62] and Gruner [17] have discussed the relations of these crystals, and the latter suggests the desirability of a re-examination of the tentative structure previously advanced for chrysotile by Warren and Bragg [73]. Caillère [6], Gruner [20], Thilo and Rogge [71], and Thilo [72] have examined the effects of heat treatment of serpentines and amphiboles, but the results appear to be somewhat inconclusive.

§4. SHEET STRUCTURES. THE MICA AND CLAY GROUPS

A detailed knowledge of this group of structures is highly important for a proper understanding of the physical properties of soil and for its application

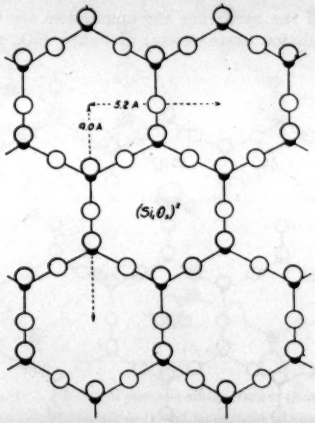


Figure 3. The hexagonal type of silicon-oxygen sheet. This sheet of tetrahedra is found in all the structures described in § 4.

(From *Atomic Structure of Minerals*, by W. L. Bragg (Cornell University Press and Oxford University Press).)

in clay technology. The group is also of special interest because of the great volume of work which is now appearing, and we shall therefore devote more space to this than to other sections.

If the process of polymerization which results in the formation of a double amphibole chain from the single pyroxene chain is continued, a tetrahedron sheet results, of composition Si_2O_5, in which every SiO_4 group shares three oxygen atoms with three other SiO_4 groups. The sheet is illustrated in figure 3, and is the fundamental unit in all mica and clay structures.

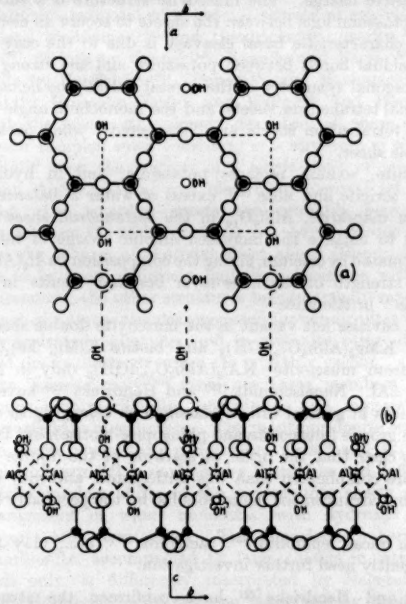


Figure 4. (*a*) The hexagonal network of linked silicon-oxygen tetrahedral groups, with hydroxyl at the centre of each ring of six vertices. (*b*) Two of the above sheets seen edgewise, with their vertices turned inwards. The layers of vertices and hydroxyls pack into each other, providing situations for Al or Mg atoms. The whole forms a firmly bound double sheet. Certain O and OH atoms which would otherwise be superimposed have been slightly displaced.

(From *Atomic Structure of Minerals*, by W. L. Bragg (Cornell University Press and Oxford University Press).)

Only one mica structure, that of muscovite $KAl_2(AlSi_3O_{10})(OH)_2$, has been completely analyzed [33], though Pauling [57] suggested in 1930 a general scheme for the various subdivisions of the mica group. In muscovite Al^{3+} replaces Si^{4+} in one tetrahedron in four, so that the tetrahedron sheet is of composition

$AlSi_3O_{10}$; two such sheets are placed together with vertices pointing inwards, aluminium atoms cross-link the vertices, and hydroxyl ions are added to complete the octahedral coordination of these aluminium atoms, as shown in figure 4. The atoms forming such a double sheet are very strongly bonded, and the whole forms a structural unit of composition $Al_2(AlSi_3O_{10})(OH)_2$, and with unit negative charge. The muscovite structure is a succession of such sheets, with potassium ions between the sheets to secure an electrically neutral crystal. The characteristic basal cleavage is due to the easy rupture of the very feeble residual bonds between potassium and the strong double sheets, the pseudohexagonal symmetry of the crystal reflects the hexagonal character of the individual tetrahedron sheets, and the monoclinic angle is fixed by the way in which tetrahedron sheets are " staggered " when packed together to form the double sheet.

In paragonite, sodium replaces potassium, and in hydromuscovite [4], glauconite [43], sericite and illite [11], excess of water is balanced by deficit of potassium. In margarite, $AlSi_3O_{10}$ in the tetrahedron sheet is replaced by $Al_2Si_2O_{10}$, and to balance the increased anionic charge of the double sheet, potassium is replaced by calcium, giving the composition $CaAl_2(Al_2Si_2O_{10})(OH)_2$; the enhanced strength of the inter-layer bonding results in a less facile cleavage in this " brittle mica "

Octahedral cavities left vacant in the muscovite double sheet are occupied in phlogopite $KMg_3(AlSi_3O_{10})(OH)_2$ and biotite $K(Mg, Fe)_3(AlSi_3O_{10})(OH)_2$ which differ from muscovite $KAl_2(AlSi_3O_{10})(OH)_2$ only in having 3Mg or 3(Mg, Fe) for 2Al. Nagelschmidt [49] and Hendricks [25] have demonstrated clearly that micas in general can be classified as belonging to one or other of these two main groups (muscovite and phlogopite-biotite), but from Hendricks' work it is also clear that the order of stacking of the double sheets is often very much more complicated than was anticipated, and may even introduce a type of random arrangement made possible by the peculiarly high symmetry of the sheets.

The lithium micas lepidolite [63], taeniolite [47], etc., may form a special group, and urgently need further investigation.

Gruner [15] and Hendricks [22] have confirmed the structural scheme originally proposed by Pauling [57] for talc $Mg_3(Si_4O_{10})(OH)_2$ and pyrophyllite $Al_2(Si_4O_{10})(OH)_2$. In both, the tetrahedra are occupied exclusively by silicon, so that the double sheets are electrically neutral, and the sheets are stacked in a partly random manner without the addition of cations such as bind the double sheets together in muscovite, biotite, etc.; hence the extreme softness of these minerals.

The chlorites, examined by Pauling [58] and McMurchy [42], are based on alternate negatively-charged mica-like sheets of composition $Mg_3(AlSi_3O_{10})(OH)_2$ to $Mg_2Al(Al_2Si_2O_{10})(OH)_2$, and positively-charged brucite-like sheets of composition $Mg_2Al(OH)_6$. Cookeite [4] is probably similar. In vermiculites [16], neutral talc-like sheets alternate with layers of water molecules arranged, according to Hendricks and Jefferson [24], in a hexagonal net by the

operation of hydroxyl bonds, in a way which explains how one half of the water content may be removed without appreciable change in the x-ray pattern [16]. The stilpnomelanes are probably similar [19], [20].

The clay structures have recently been reviewed by Sir William Bragg [3] and by Bannister [1], and a special section of *Mineralogical Abstracts* is now [46] devoted to these materials.

Following Pauling's proposal of a general structural scheme [58], kaolinite has been examined by Gruner [12] and Hendricks [21], nacrite by Gruner [14] and Hendricks [26], dickite by Gruner [13], Ksanda and Barth [37] and Hendricks [23], anauxite by Hendricks [21], Gruner [18], and Machatschki [41], halloysite by Mehmel [45], Hofmann Endell and Wilm [29] and Hendricks [23], [24]. By contrast with the micas, etc., already described, all these structures are based upon a *polar* complex sheet consisting of a structure of aluminium and hydroxyl ions built upon the vertices of a *single* sheet of tetrahedra, with ideal composition $Al_2Si_2O_5(OH)_4$, and electrically neutral (cf. pyrophyllite $Al_2(Si_4O_{10})(OH)_2$). The complex sheets are probably held together by hydroxyl bonds between the hydroxyl layer on the top of one sheet and the oxygen layer on the bottom of the next sheet above (Hendricks [26]). Only nacrite appears to be quite definitely built in accordance with space-group symmetry requirements, the other structures being partially random (cf. talc). The hexagonal net of water molecules suggested for vermiculites [24] is probably present in fully-hydrated halloysite, and in anauxite some aluminium and hydroxyl sites of the kaolinite structure are probably vacant.

According to Nagelschmidt [50] there are three members of the montmorillonite group—montmorillonite, nontronite, and magnesium-beidellite, which differ very slightly in structure. The montmorillonite structure, based on pyrophyllite sheets, suggested by Hofmann, Endell and Wilm [28] and by Maegdefrau and Hofmann [44], is modified by Hendricks and Jefferson [24], who introduce partially random shifts of the pyrophyllite sheets and suggest a hexagonal arrangement of water molecules, with hydroxyl bonds, as in hydrated halloysite and vermiculites. The remarkable reversible dehydration of montmorillonite, accompanied by large change in lattice-dimensions along the *c*-axis only, is differently interpreted by Nagelschmidt [48] and Maegdefrau and Hofmann [44], on the one hand, and by Bradley, Grim and Clark [2] on the other.

This group of structures is of the utmost importance both by reason of its intrinsic interest and because of the widespread practical application of the materials.

§5. FRAMEWORK STRUCTURES

The various forms of silica are framework structures in which every SiO_4 tetrahedron shares all its corners with other tetrahedra ("Minerals", chapter VII); recent work has revealed details of the structures of α-quartz [74], α-cristobalite [53], and β-cristobalite [54], which are of great interest for comparison with the β (high temperature) structures of quartz and tridymite.

When some tetrahedra in a framework contain Al^{3+} instead of Si^{4+}, cations must be added to maintain electrical neutrality. Zeolites and ultramarines are structures of this kind ("Minerals", chapter XVI). Leucite, $KAlSi_2O_6$, well known as an example of pseudo-symmetry due to twinning, is probably structurally similar to analcite $NaAlSi_2O_6H_2O$, but Wyart [75] and Onorato [56] are not in full agreement as to the true symmetry of the unit cell, and no structure determination has been made. Pollucite, also similar to analcite, has been examined by Strunz [65], Richmond and Gonyer [61], and Náray-Szabó [52]; details of new work on its dehydration properties by Fleischer and Ksanda [9] are not yet published. Podschus, Hofmann, and Leschewski [60] have refined the ideal structure of ultramarine, originally analyzed by Jaeger [34].

The felspars are also framework structures ("Minerals", chapter XV). New work by Ito and Inuzuka [31], Ito [32], Chao, Smare and Taylor [7], and Chao and Taylor [8], on the potash-soda lamellar felspars encourages the hope that some light may be thrown on the relations between structural detail and physical properties of these fascinating but complex minerals.

§6. CONCLUSION

The writer of this section has thought it best, in the very limited space available, to assume that the reader is either acquainted with the general nature of crystal structures or has access to the standard work "Minerals" to which frequent reference has been made, and he has attempted to show in what directions structure research on silicates is most active at the present time. It is, unfortunately, not possible to include any account of the more technical work on ceramics, refractories, glasses (including borates), soils, and so on; for these the appropriate journals should be consulted.

REFERENCES

(1) Bannister, F. A. *Ann. Rep. Chem. Soc.* **35**, 187 (1939).
(2) Bradley, W. F., Grim, R. E. and Clark, G. L. *Z. Kristallogr.* **97**, 216 (1937).
(3) Bragg, Sir W. H. *Proc. Roy. Inst.* **30**, 39 (1938).
(4) Brammall, A., Leech, J. G. C. and Bannister, F. A. *Miner. Mag.* **24**, 507 (1937).
(5) Cagliotti, V. *Atti V Congr. Naz. Chim.* **1**, 310 (1936).
(6) Caillère, S. *Bull. Soc. Franc. Miner.* **59**, 163 (1936).
(7) Chao, S. H., Smare, D. L. and Taylor, W. H. *Miner. Mag.* **25**, 338 (1939).
(8) Chao, S. H. and Taylor, W. H. *Nature, Lond.*, **143**, 1024 (1939).
(9) Fleischer, M. and Ksanda, C. J. *Amer. Min.* **24**, 185 (1939).
(10) Goldschmidt, V. M. *Nachr. Ges. Wiss. Göttingen*, 184 (1931).
(11) Grim, R. E., Bray, R. H., and Bradley, W. F. *Amer. Min.* **22**, 813 (1937).
(12) Gruner, J. W. *Z. Kristallogr.* **83**, 75 (1932).
(13) Gruner, J. W. *Z. Kristallogr.* **83**, 394 (1932).
(14) Gruner, J. W. *Z. Kristallogr.* **85**, 345 (1933).
(15) Gruner, J. W. *Z. Kristallogr.* **88**, 412 (1934).

(16) Gruner, J. W. *Amer. Min.* **19**, 557 (1934).
(17) Gruner, J. W. *Amer. Min.* **22**, 97 (1937).
(18) Gruner, J. W. *Amer. Min.* **22**, 855 (1937).
(19) Gruner, J. W. *Amer. Min.* **22**, 912 (1937).
(20) Gruner, J. W. *Amer. Min.* **23**, abstracts (7) (1938).
(21) Hendricks, S. B. *Z. Kristallogr.* **95**, 247 (1936).
(22) Hendricks, S. B. *Z. Kristallogr.* **99**, 264 (1938).
(23) Hendricks, S. B. *Amer. Min.* **23**, 295 (1938).
(24) Hendricks, S. B. and Jefferson, M. E. *Amer. Min.* **23**, 851 and 863 (1938).
(25) Hendricks, S. B. *Nature, Lond.*, **143**, 800 (1939).
(26) Hendricks, S. B. *Z. Kristallogr.* **100**, 509 (1939).
(27) Heritsch, H. *Z. Kristallogr.* **98**, 351 (1937–38) and **99**, 466 (1938).
(28) Hofmann, U., Endell, K. and Wilm, D. *Z. Kristallogr.* **86**, 340 (1933).
(29) Hofmann, U., Endell, K. and Wilm, D. *Z. angew. Chem.* **47**, 539 (1934).
(30) Hutton, C. O. and Fankuchen, I. *Miner. Mag.* **25**, 172 (1938).
(31) Ito, T. and Inuzuka, H. *Z. Kristallogr.* **95**, 404 (1936).
(32) Ito, T. *Z. Kristallogr.* **100**, 297 (1938).
(33) Jackson, W. W. and West, J. *Z. Kristallogr.* **76**, 211 (1930) and **85**, 16 (1933).
(34) Jaeger, F. M. *Trans. Faraday Soc.* **25**, 320 (1929).
(35) Kokkoros, P. *Z. Kristallogr.* **96**, 417 (1937).
(36) Kokkoros, P. *Z. Kristallogr.* **99**, 38 (1938).
(37) Ksanda, C. J. and Barth, T. F. W. *Amer. Min.* **20**, 631 (1935).
(38) Laves, F. *Z. Kristallogr.* **73**, 202 and 275 (1930).
(39) Levi, G. R. and Peyronel, G. *Z. Kristallogr.* **92**, 190 (1936); Peyronel, G., *Z. Kristallogr.* **94**, 311 (1936).
(40) Machatschki, F. *Zbl. Min. Geol. Paläont.* **1928** A, 97.
(41) Machatschki, F. *Amer. Min.* **23**, 117 (1938).
(42) McMurchy, R. C. *Z. Kristallogr.* **88**, 420 (1934).
(43) Maegdefrau, E. and Hofmann, U. *Z. Kristallogr.* **98**, 31 (1937).
(44) Maegdefrau, E. and Hofmann, U. *Z. Kristallogr.* **98**, 299 (1937).
(45) Mehmel, M. *Z. Kristallogr.* **90**, 35 (1935); *Chem. d. Erde*, **11**, 1 (1937); *Fortschr. Min.* **21**, 80 (1937)
(46) *Min. Abstr.* **7**, 94 (1938)
(47) Miser, H. D. and Stevens, R. E. *Amer. Min.* **23**, 104 (1938).
(48) Nagelschmidt, G. *Z. Kristallogr.* **93**, 481 (1936).
(49) Nagelschmidt, G. *Z. Kristallogr.* **97**, 514 (1937).
(50) Nagelschmidt, G. *Miner. Mag.* **25**, 140 (1938).
(51) Náray-Szabó, St. *Z. phys. Chem.*, B, **9**, 356 (1930).
(52) Náray-Szabó, St. *Z. Kristallogr.* **99**, 277 and 291 (1938). See also Taylor, W. H., *Z. Kristallogr.* **99**, 283.
(53) Nieuwenkamp, W. *Z. Kristallogr.* **92**, 82 (1935)
(54) Nieuwenkamp, W. *Z. Kristallogr.* **96**, 454 (1937).
(55) Niggli, P. *Z. Kristallogr.* **86**, 121 (1933).
(56) Onorato, E. *Period. Miner.* **9**, 85 (1938).
(57) Pauling, L. *Proc. Nat. Acad. Sci., Wash.*, **16**, 123 (1930).
(58) Pauling, L. *Proc. Nat. Acad. Sci., Wash.*, **16**, 578 (1930).
(59) Pauling, L. and Sherman, J. *Z. Kristallogr.* **96**, 481 (1937).
(60) Podschus, E., Hofmann, U. and Leschewski, K. *Z. anorg. Chem.* **228**, 305 (1936).
(61) Richmond, W. E. and Gonyer, F. A. *Amer. Min.* **23**, 783 (1938).
(62) Selfridge, G. C. *Amer. Min.* **21**, 463 (1936).
(63) Stevens, R. E. *Amer. Min.* **23**, 607 (1938).
(64) Strunz, H. *Z. Kristallogr.* **94**, 60 (1936).

(65) Strunz, H. *Z. Kristallogr.* **95**, 1 (1936).
(66) Strunz, H. *Z. Kristallogr.* **96**, 7 (1937).
(67) Strunz, H. *Z. Kristallogr.* **98**, 60 (1937–38).
(68) Strunz, H. *Z. Kristallogr.* **99**, 513 (1938).
(69) Taylor, W. H. and West, J. *Z. Kristallogr.* **70**, 461 (1929).
(70) Taylor, W. H. *Z. Kristallogr.* **68**, 503 (1928).
Taylor, W. H. and Hey, J. S. *Z. Kristallogr.* **80**, 428 (1931).
Taylor, W. H. and Thornely, P. C. *Mem. Manchr. Lit. Phil. Soc.* **83**, 17 (1939).
(71) Thilo, E. and Rogge, G. *Ber. dtsch. chem. Ges.* **72**, 341 (1939).
(72) Thilo, E. *Z. Kristallogr.* **101**, 345 (1939).
(73) Warren, B. E. and Bragg, W. L. *Z. Kristallogr.* **76**, 201 (1930).
(74) Wei, P'ei-Hsui. *Z. Kristallogr.* **92**, 355 (1935).
(75) Wyart, J. *C.R. Acad. Sci., Paris*, **203**, 938 (1936).
(76) Zachariasen, W. H. *Z. Kristallogr.* **73**, 1 (1930).
(77) Zachariasen, W. H. *Z. Kristallogr.* **73**, 7 (1930).

III. THE EXAMINATION OF TERNARY ALLOY SYSTEMS BY X RAYS

By H. LIPSON, D.Sc.
Cavendish Laboratory, Cambridge

§1. INTRODUCTION

X-ray methods have been applied with conspicuous success to the study of binary alloy systems, and their extension to ternary systems seems to offer still greater potentialities. In binary systems one can trace the variation of any one property within a single-phase field, and a large amount of such data has been collected. In ternary systems one can observe the modification of the variation by the addition of the third component, and thus gain valuable evidence concerning its cause. For example, the addition of nickel to the copper-aluminium alloys [6] shows that the phase boundaries of the γ-structures [4] follow the directions to be expected from Konobeevsky's theory [14], and are not directly dependent on atomic sizes.

The pioneer work, therefore, consists in mapping out the extent of the various phase-fields, in order to see in what systems such information may be provided.

The experimental difficulties involved in the exploration of a complete ternary diagram are not so great as a merely numerical survey of the problem might indicate. The number of alloys necessary for the study of a binary system of average complexity is of the order of 50 [8]. To cover a ternary diagram with the same degree of detail would necessitate about 1000 alloys. However, if the three binary systems are known, and if the validity of the geometrical rules concerning ternary equilibrium may be assumed, the number of alloys may be reduced to the order of 100 [9].

§2. FORM OF A TERNARY DIAGRAM

One of the properties of an equilateral triangle is that the sum of the perpendiculars from any point within it on to the three sides is constant. Such a triangle may thus be used for plotting the compositions of ternary alloys [17]. The corners represent the component metals, the edges represent the binary systems, and the lengths of the perpendiculars from any point on to the sides give the proportions of the components in the alloy represented by that point. This method of representation is shown in figure 1.

The general form of the equilibrium diagram may best be explained in terms of free energy, although, since this cannot yet be evaluated as a function of the composition, the method cannot do more than illustrate the principles of equilibrium. The fact that the free energy of a system in equilibrium must be a minimum was first applied by Roozeboom [15] to binary systems. The principle of his method is shown in figure 2. The curves are those of free

energy against composition, for particular structures, α, β, and γ, and the limits of extent of the structures are given by the points of contact of the common tangents to neighbouring curves.

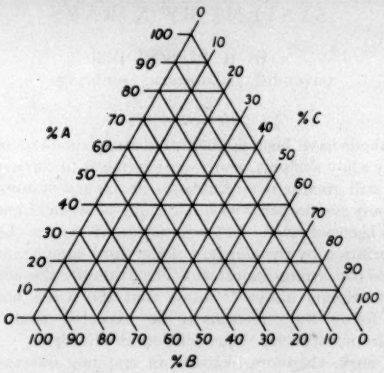


Figure 1. Representation of compositions of ternary alloys.

In ternary systems the free-energy curves are replaced by surfaces, and the tangent line by a tangent plane. This is illustrated in figure 3, in which is shown a ternary system in which each of the components, A, B, C, has a

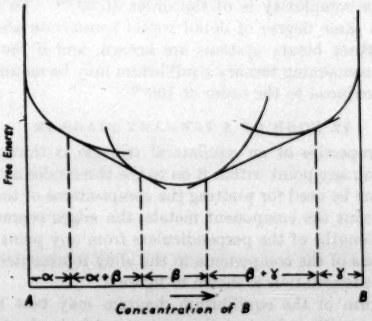


Figure 2. Derivation of phase boundaries from free-energy curves in a binary system.

different structure, α, β, γ, respectively. Suppose that the surfaces are convex downwards, and that the common tangent plane touches them in the points E, H, M. Then the triangle of which these points are the corners is

an area of three-phase equilibrium. Suppose, now, that the tangent plane be tilted so that it loses contact with the point M. It will thus have one degree of freedom, and as it moves the points of contact with the surfaces α and β will trace out the curves ED and HK. These curves represent limits of extent

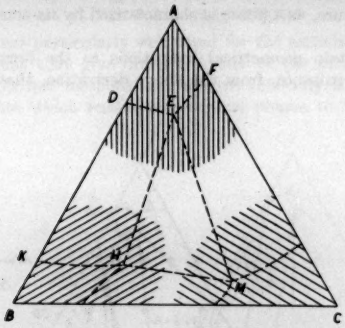


Figure 3. Derivation of phase boundaries from free-energy surfaces in a ternary system.

of the structures α and β, and between them is a region of two-phase equilibrium, $\alpha+\beta$. The points at which the tangent plane, in any given position, touches the two curves represent phases which can coexist in equilibrium

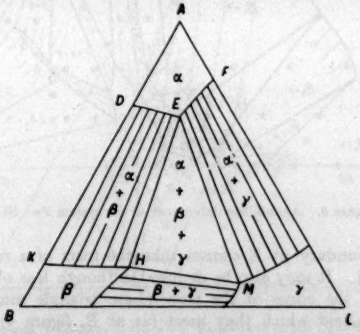


Figure 4. Typical isothermal ternary diagram, derived from figure 3.

(conjugate phases), and the lines joining them are called "tie-lines" [17] or "co-nodal lines" [18]. In figure 4 are shown the details of the phase diagram produced in this way. These details will change with temperature, as the free-energy surfaces change relatively to each other, but the form of the diagram

will remain the same, until, for example, another free-energy surface (perhaps that of the liquid) intersects the plane through the points E, H, M.

Areas such as ADEF are regions of single phase, or, more concisely, "single-phase fields". It is incorrect to call such regions "phases", as is sometimes done in the literature, as a phase is characterized by its composition as well as by its structure.

There are certain geometrical limitations to the form of the diagram. It is obvious, for instance, from the above derivation, that the discontinuity

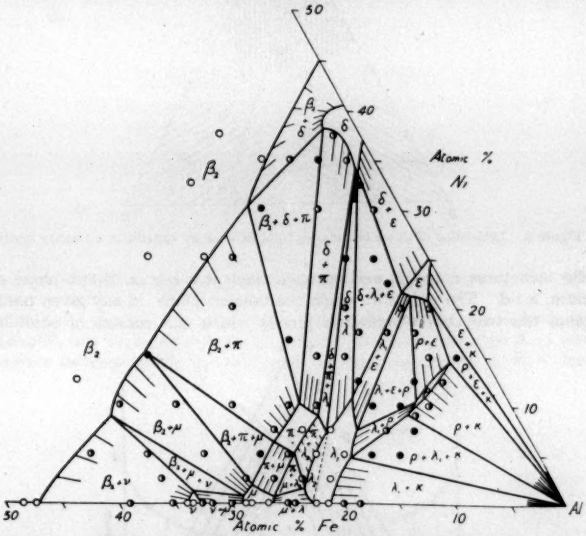


Figure 5. Aluminium-rich corner of the system Fe—Ni—Al.

on the phase boundary at E cannot take the form of a re-entrant into the single-phase field. It may also be deduced, although less obviously, that the continuations of the edges of the three-phase triangle must lie either both inside the phase-field which they meet (as at E, figure 4), or both outside (as at H) [16].

Actual phase diagrams may be more or less complicated than this particular illustration. They may be less complicated if the three components, or two of them, have similar structures, although this does not follow. They will be more complicated if one or more intermediate structure exists in any of the binary systems, and particularly if a ternary structure appears. In the general

case, the diagram will be covered by a patchwork of single-, two-, and three-phase fields, as is beautifully exemplified by the aluminium-rich corner of the Fe—Ni—Al diagram [9] (figure 5).

§3. VALUE OF X-RAY METHODS

X-ray methods are particularly well fitted for the examination of ternary alloy systems, in that they give direct evidence of the phases which are in equilibrium. Microscopic methods can achieve this only if it is possible to find etching reagents which enable the various phases to be identified, for

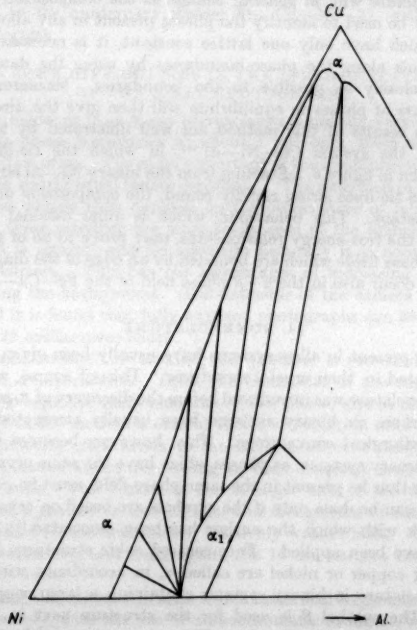


Figure 6. Directions of tie-lines in the α+α phase-field of the Cu—Ni—Al system.

instance, by their colours, or if the phases themselves have distinctive crystalline appearances [10]. With X rays, however, each phase usually gives a characteristic pattern, and the problem of recognizing the phases present resolves itself into that of detecting the diffraction patterns of possible structures. On this account, the method is not absolute ; that is, one cannot tell, from the

photograph of a given alloy, of how many phases it is composed. It is necessary first for the photographs of all possible structures to be obtained, and, if a ternary intermetallic compound is involved, it must be proved that its photograph *is* that of a single phase. This can be done only by the accumulation of evidence from several alloys. For this reason, it must be emphasized that X-ray evidence of the phase-structure of any single alloy is probably less reliable than the microscopic evidence.

A fundamental advantage of X-ray methods, however, is the possibility of determining the directions of tie-lines [2]. The lattice constants of any particular structure will, in general, change as the composition changes, and thus they may be used to identify the phases present in any alloy. For cubic structures, which have only one lattice constant, it is necessary to find the variation of this along the phase-boundaries by using the data from alloys made up as closely as possible to the boundaries. Measurements of the lattice constants of phases in equilibrium will then give the directions of the tie-lines. The results of this method are well illustrated by the two-phase field $\alpha+\alpha_1$ in the system Cu—Ni—Al [6], in which the tie-lines have the directions shown in figure 6. Starting from the binary Ni—Al edge of the two-phase field, the tie-lines swing rapidly round, the composition of the α_1 phase remaining constant. This behaviour, which is quite rational if considered in the light of the free-energy relationships, may prove to be of general occurrence in two-phase fields which are bounded by an edge of the diagram. It has been found to occur also in the $\alpha+\beta$ phase field of the Fe—Cu—Al system [3].

§4. NOMENCLATURE

The phases present in alloy systems have usually been given conventional symbols unrelated to their crystal structures. This, of course, was inevitable when the nomenclature was introduced before the discovery of X-ray diffraction, and X-ray workers on binary systems have usually attempted to support the older metallurgical conventions. This, however, becomes impossible in dealing with ternary systems, as phases which have the same crystal structure, and which may thus be present in the same phase-field, must be given the same symbol. This can be done only if the symbols are based on crystal structure.

In the work with which the author has been associated [5] the following conventions have been applied. Face-centred cubic structures in the phase-fields including copper or nickel are called α, in accordance with the metallurgical nomenclature in binary systems containing a large copper-rich solid solution [12]. The symbol β is used for the structure next in frequency of occurrence—the body-centred cube. This also agrees with the generally accepted nomenclature in the systems in which this structure follows the α-solid solution, but, unfortunately, it clashes with what is perhaps the best established symbol in metallurgy, α-iron. Nevertheless, it is impossible to evade the issue. The symbol γ is reserved for structures similar to that of γ-brass, but other structures are considered to be of insufficient practical or theoretical importance for special symbols to be assigned to them.

Superlattices are distinguished by suffixes. Thus, a face-centred structure in which the atoms at the corners are different from those at the centres of the faces is called α_1. It is quite possible for an α phase-field to include an α_1 region, and thus it is inadvisable to give the superlattice structure an entirely different symbol [1]. The body-centred cubic structure of the type Fe_3Al is called β_1, and that of the type FeAl, β_2. When two structures of the same type are in equilibrium, they must be differentiated for the purpose of reference. This is done by adding a superscript to one of the symbols, as, for example, in the phase-field $\alpha_1+\alpha_1'$ in the system Fe—Ni—Al [9].

It has been found possible to apply these conventions to several ternary systems, and to the quaternary system Fe—Cu—Ni—Al, without any confusion or ambiguity.

§5. SOME DIFFICULTIES IN THE INTERPRETATION OF THE X-RAY PHOTOGRAPHS

While the evidence from X-ray photographs is usually very clear, enabling one to fix the phase boundaries to about 1 per cent, there are cases where special difficulties arise. The most important is the equilibrium of two similar structures with nearly equal lattice constants. The detection of small amounts of one of these phases then depends on the observation of faint satellites to the spectra; these satellites are more easily seen at the higher orders, where the resolution is greater, and for adequate accuracy it is necessary to use a rather large camera. This has the advantages of increasing the resolution and of reducing the background. The diameter of the camera normally used is 19 cm., and it is found that fully-exposed photographs can be obtained with exposures of 25 milliampere-hours.

An additional method of tackling the problem is provided if one of the structures has a superlattice. Then, the intensities of the superlattice lines may be used to estimate small amounts of this phase, and in order to increase the sensitivity of the method a careful choice of radiation must be made [7].

Another difficulty often arises, in the interpretation of photographs of two or three complicated structures, owing to their frequent resemblance to the body-centred cube. This results in the stronger lines of the different phases overlapping to a greater or less extent, and recognition of subsidiary phases is thus dependent on the detection of relatively weaker lines. This, of course, reduces the sensitivity, and here, again, the 19-cm. camera has proved to be of considerable value.

Nevertheless, difficulties still arise which make still larger cameras desirable, combined, possibly, with the use of monochromatic radiation reflected from a crystal [11]. Also, in certain cases [4], purely monochromatic radiation is necessary, the presence of the $\alpha_1\alpha_2$ doublets being a definite hindrance to the correct interpretation of the fine structure of the photographs.

Another direction in which development is necessary is the study of the reactions which occur in the solid state. This involves the use of a camera in which the specimen may be held at a high temperature [13], or in which

photographs may be taken while the specimen is heating or cooling. This is necessary if x rays are to make their full contribution to the solution of the problems which confront the physicist and metallurgist.

REFERENCES

(1) Alexander and Vaughan. *J. Inst. Met.* **61**, 247 (1937).
(2) Anderson and Jette. *Trans. Amer. Soc. Metals*, **24** (2), 375 (1936).
(3) Bradley and Goldschmidt. *J. Inst. Met.* **65** (advance copy) (1939).
(4) Bradley, Goldschmidt and Lipson. *J. Inst. Met.* **63**, 149 (1939).
(5) Bradley, Goldschmidt, Lipson and Taylor. *Nature, Lond.*, **140**, 543 (1937).
(6) Bradley and Lipson. *Proc. Roy. Soc.* A, **167**, 421 (1938).
(7) Bradley and Rodgers. *Proc. Roy. Soc.* A, **144**, 340 (1934).
(8) Bradley and Taylor. *Proc. Roy. Soc.* A, **159**, 56 (1937).
(9) Bradley and Taylor. *Proc. Roy. Soc.* A, **166**, 353 (1938).
(10) Desch. *Metallography*, p. 314. (London, 1937.)
(11) Fankuchen. *Nature, Lond.*, **139**, 193 (1937).
(12) Hansen. *Der Aufbau der Zweistofflegierungen*. (Berlin, 1936.)
(13) Hume-Rothery and Reynolds. *Proc. Roy. Soc.* A, **167**, 25 (1938).
(14) Konobeevsky. *J. Inst. Met.* **63**, 161 (1939).
(15) Roozeboom. *Z. phys. Chem.* **30**, 385 (1899).
(16) Roozeboom. *Die Heterogenen Gleichgewichte*, II. (Brunswick, 1901.)
(17) Stokes. *Proc. Roy. Soc.* A, **49**, 147 (1891).
(18) Vogel. *Handbuch der Metallphysik*, II: *Die Heterogenen Gleichgewichte*. (Leipzig, 1937.)

IV. ORDER-DISORDER PHENOMENA IN ALLOYS

By E. T. GOODWIN, M.A., Ph.D.
The University, Sheffield

§ 1. INTRODUCTION

Many alloys, for suitable concentrations of their components, form single-phase systems having characteristic lattice structures. If the metal atoms are distributed at random among the possible atomic sites then on the average these sites will be indistinguishable, and an X-ray photograph will show the lines characteristic of the lattice just as if it were a pure metal. If, however, there are preferential sites for certain of the atoms they may distribute themselves regularly throughout the lattice, thus superimposing a pattern upon it and forming a *superlattice*. Correspondingly *superlattice lines* appear in the X-ray diffraction pattern, since some parallel planes, previously indistinguishable, now contain different distributions of atoms; they have, therefore, different reflection coefficients, and the corresponding reflected waves that are exactly out of phase no longer cancel. A typical example is Cu_3Au, first investigated by Johansson and Linde [13]. The copper and gold atoms were at first found to be distributed at random on the sites of a face-centred cubic structure, but after prolonged annealing, analysis indicated that the gold atoms occupied the cube corners and the copper atoms the face centres.

This ordering of the atoms among the available sites manifests itself in several other ways, such as anomalies in the specific heat and resistivity below a certain critical temperature. Explanations of these order-disorder phenomena first put forward by Gorsky [12], Borelius [2], Dehlinger [8] and others were subsequently considerably extended by Borelius [3]. His theory, though of great generality, is perforce of a somewhat empirical nature, and it is on the theories of Bragg and Williams [6], and particularly of Bethe [1], that most recent work has been based.

In this article it is not possible to give an account of the great volume of research, both practical and theoretical, carried out on order-disorder transformations in alloys. For this the reader is referred to the report in *Reviews of Modern Physics* [20], which contains an excellent bibliography, and to various papers in the Physical Society *Conference on conduction in solids* [22]. Here it seems most desirable to outline the methods of Bragg and Williams and of Bethe, and then to give some indication of the extent to which the experimental results are thus predicted, and to what causes discrepancies may be attributed.

§ 2. THEORETICAL DETERMINATION OF THE ORDER

How may the order be expected to depend on the temperature? In general the ordered state will be of less energy than a disordered one and thus more

likely to occur, but, on the other hand, there are far more disordered than ordered states. At low temperatures the energy effect is the more important and the atoms of the alloy should take up positions of high order ; at high temperatures the superabundance of disordered states should be the decisive factor causing the adoption of a state of at most partial order. This behaviour is, indeed, frequently observed and will be seen to be successfully predicted by the theories developed below.

The method of Bragg and Williams. Bragg and Williams consider a single-phase system wherein atoms of two types, A and B, can be rearranged. Let there be rn A atoms and $(1-r)n$ B atoms (the names so chosen that $r \leqslant \frac{1}{2}$) occupying, in a state of complete order, lattice points which will be called α- and β-sites respectively. n is not necessarily equal to the total number N of atoms in the lattice, since some of these may take no part in the order-disorder transformation ; for example, in Fe_3Al the cube corners are always occupied by iron atoms. The degree of order S must clearly depend on the fraction p of A atoms actually occupying α-sites at any instant, and Bragg and Williams accordingly define S by the relation

$$S = \frac{p-r}{1-r}. \qquad \ldots.(1)$$

$S=1$ corresponds to $p=1$ and hence to perfect order ; $S=0$ to $p=r$, i.e. the atoms are distributed at random, which is complete disorder.

To investigate the dependence of S on the temperature T it is now supposed that when an A-atom on an α-site and a B-atom on a β-site are interchanged, thus producing a pair of disordered atoms, there is an energy increase of V. Then by Boltzmann's relation

$$\frac{\text{Number of ordered A–B pairs}}{\text{Number of disordered A–B pairs}} = e^{V/kT}. \qquad \ldots.(2)$$

The left-hand side is expressible in terms of p and hence of S. The energy of interchange V will be a function of S increasing steadily from zero when the crystal is completely disordered ($S=0$) to a maximum V_0 when the order is perfect ($S=1$). Bragg and Williams accordingly assume

$$V = V_0 S, \qquad \ldots.(3)$$

which in conjunction with (2) yields an equation whose root S is the degree of order appropriate to the given temperature T.

For a simple AB-type alloy, $r=\frac{1}{2}$ and the equation reduces to

$$S = \tanh SV_0/4kT. \qquad \ldots.(4)$$

The reader will note the similarity between this equation and that of the Weiss-Heisenberg theory of ferromagnetism. Analogously to the Curie temperature there is here a critical temperature $T_c = V_0/4k$. For $T>T_c$ there is only one root $S=0$, i.e. the alloy is disordered, but for $T<T_c$ there is also a non-zero root, shown in figure 1. This state of partial order has less free energy than the disordered state and so it is the equilibrium state at a temperature $T<T_c$. As $T \to T_c$, S vanishes with a vertical tangent ; the physical explanation of this

is that the less the degree of order the easier is the subsequent disordering, which finally becomes catastrophic.

If $r \neq \frac{1}{2}$ the behaviour of the roots of (4) is somewhat different.* As T decreases there is at first only one root, $S=0$, but later two non-zero roots appear. Of these the smaller represents a state which is always unstable, while the larger corresponds to a free energy at first greater than that of the disordered state. The critical temperature is that at which these free energies become equal, and below which the free energy of the state of finite order is the lower. Thus it lies below the temperature at which non-zero roots first appear, and at it the degree of order will jump discontinuously to a finite value and then steadily increase to unity at the absolute zero (see figure 1).

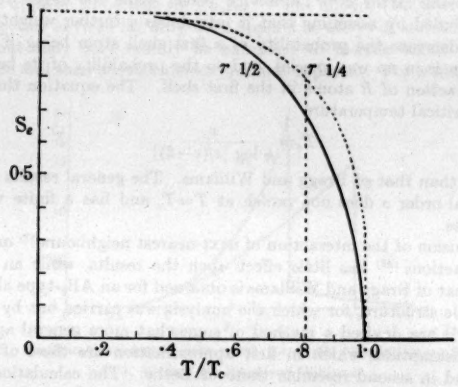


Figure 1. (Reproduced by courtesy of Professor W. L. Bragg.)

The method of Bethe. The chief weakness of the Bragg and Williams theory lies in the fact that, whatever the temperature, a certain amount of local ordering among near neighbours will be present. This should not greatly influence the formation of superlattice lines, or the resistivity, but might be expected to be of importance in considerations of energy and specific heat which depend essentially on near neighbours. Bethe accordingly defines the degree of local order σ as the difference between the probabilities of finding an unlike and a like pair as nearest neighbours.

The case considered is an AB-type alloy of N atoms each having z nearest neighbours. If the crystal contains at any instant n_{AB} pairs of nearest neighbours (AB), the interaction energy being v_{AB}, and similarly, then the energy

* For a comprehensive account of this and of the whole theory, based on statistical mechanics, see Fowler's *Statistical Mechanics* [11].

of ordering is, neglecting further neighbours,

$$\left.\begin{aligned} E &= n_{AA}v_{AA}+n_{BB}v_{BB}+n_{AB}v_{AB}, \\ &= \text{const.}-n_{AB}v, \end{aligned}\right\} \quad \ldots\ldots(5)$$

where $v=\frac{1}{2}(v_{AA}+v_{BB})-v_{AB}$,

since $n_{AA}+\frac{1}{2}n_{AB}=n_{BB}+\frac{1}{2}n_{AB}=\frac{1}{4}Nz$.

The energy v, which must be positive for ordering to take place, is connected with the Bragg V_0 by the relation $V_0=2zv$.

Bethe now takes any atom in the lattice and considers the probabilities of its being right (R), i.e. A on α- or B on β-site, or wrong (W). If it is R the probability of an atom in the first shell of neighbours being W is weighted by a Boltzmann factor $e^{-v/kT}$, and vice versa, while the effect of long-range order is included by assuming that it introduces a further weighting factor ϵ tending to decrease the probability of a first shell atom being W. Now the central atom is in no way special, and so the probability of its being R must equal the fraction of R atoms in the first shell. The equation thus obtained leads to a critical temperature

$$T_c=\frac{v}{k\log\left[z/(z-2)\right]}$$

lying lower than that of Bragg and Williams. The general results are similar, but the local order σ does not vanish at $T=T_c$ and has a finite value at all temperatures.

The inclusion of the interaction of next-nearest neighbours [7] or even long-range interactions [26] has little effect upon the results, while an S–T curve similar to that of Bragg and Williams is obtained for an AB_3-type alloy of face-centred cubic structure, for which the analysis was carried out by Peierls [21]. Kirkwood [15] has devised a method of somewhat more general applicability, involving assumptions which in first approximation are those of Bragg and Williams and in second resemble those of Bethe. The calculations are very heavy and yield results similar to those of Bethe.

The sudden onset of order predicted by all three theories is amply confirmed by X-ray analysis, from the original work of Johansson and Linde [13] to that on the β-brass superstructure by Jones and Sykes [14]. The latter is outstanding in view of the similar scattering powers of Cu and Zn. These, however, show a marked difference for Zn Kα radiation which happens to lie near the Cu and Zn absorption edges, and so gives rise to observable superlattice lines. This technique of careful choice of the wave-length of the incident radiation had previously been employed by Bradley and Rodgers [5] in showing the Heusler alloys to be ordered in the ferromagnetic state.

§ 3. THE EFFECT OF ORDER ON THE SPECIFIC HEAT

Apart from the thermal energy, an alloy has a configurational energy varying with its state of order. Measurements of this energy content and the corresponding "anomalous" specific heat form a convenient way of testing the theories quantitatively, and they are found to be far from satisfactory.

The energy on the Bragg and Williams theory is

$$E=\tfrac{1}{2}nr(1-r)(1-S^2) \qquad \ldots.(6)$$

and on the Bethe theory is

$$E=\tfrac{1}{2}Nzv(1-\sigma), \qquad \ldots.(7)$$

the specific heat being dE/dT in each case.

The theoretical type of variation of C_v with temperature for an AB alloy is depicted in figure 2, the Kirkwood curve being very similar to that of Bethe. The effect of local order is seen in the finite specific heat above T_c. This is confirmed by experiment [18], [23] which, however, shows a rise to a much higher maximum followed by a very sharp but continuous fall. Theory also predicts a latent heat if $r\neq\frac{1}{2}$, corresponding to the discontinuity in S, but not for $r=\frac{1}{2}$; this appears to be true for Cu_3Au [25] but not for Cu_3Pd or MgCd [22].

The theories have so far only been applied to alloys of stoichiometric composition, but they may readily be extended to include small variations in composition and then another defect becomes apparent. For while experiments

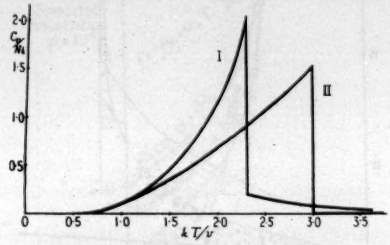


Figure 2. Configurational specific heat curves for simple cubic lattice ($z=6$) with $r=\frac{1}{4}$.
I. Bethe. II. Bragg and Williams.

have shown T_c to be a maximum for the 3 : 1 ratio of the components of an AB_3 alloy, theory predicts higher values on the 1 : 1 side [9].

The maximum specific heat could be raised if the ordering energy increased with S more rapidly than is assumed by Bragg and Williams or than is implicit in the Bethe and Kirkwood theories, and Mott [19] gives reasons for expecting this to be so. This, however, would not remove the discontinuity. Thus, while the theory seems on the right lines, the weakness appears not to lie entirely in mathematical approximations. This situation has been considerably improved by the recent calculations of Eisenschitz [10] for the specific heat of β-brass. Eisenschitz points out that the Kirkwood theory in first approximation, i.e. effectively the assumptions of Bragg and Williams, leads to a continuous specific-heat curve if applied to a lattice containing a finite number of atoms j, which is not more arbitrary than applying it to an infinite lattice. Further, it is C_p and not C_v that is measured experimentally, and there may be considerable

difference between them in the neighbourhood of T_c owing to volume changes likely to be associated with the disordering.

To account for this it is assumed that there is some volume for which the energy v_{AA} has a minimum and that v_{AA} varies as the square of the difference from this volume, similar assumptions being made for v_{BB} and v_{AB}. The theory then developed (on the lines of Bragg and Williams) contains various parameters which can be adjusted to give good agreement with experiment (figure 3).

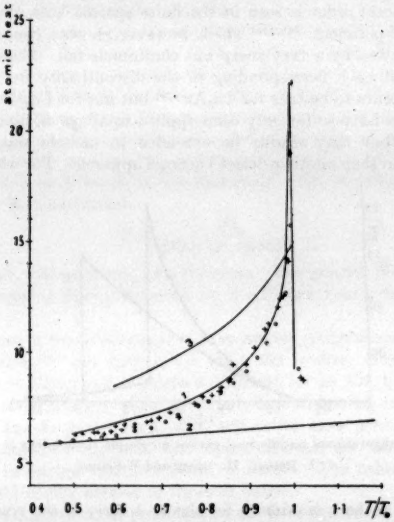


Figure 3. 1, Theoretical C_p curve. 2, specific heat of vibrations. 3, C_v curve according to Kirkwood. +, measurements by Syke. ○, measurements by Moser. (Reproduced by courtesy of Dr. Eisenschitz.)

An independent check on the parameters may be formulated and it is found that the choice is reasonable. In particular if the crystal is divided into cubes, and physical quantities are considered as derived from the cubes taken independently, together with a contribution from the effect of coupling between them, then this coupling is shown to become unimportant when the cubes contain more than 10^4 atoms each, which is the same order of magnitude as the value chosen for the parameter j in order to obtain agreement with experiment. The calculations also give a finite slope for the C_v curve, and this absence of discontinuity is shown to be necessary on general thermodynamical grounds.

§ 4. THE EFFECT OF ORDER ON THE RESISTIVITY

When $S=1$ the resistivity ρ_1 of an alloy should behave like that of a pure metal

$$\rho_1=\alpha_1 T.$$

The resistivity in the state of complete disorder, however, will contain a constant contribution due to the disordering and have the form

$$\rho_0=\rho_d+\alpha_0 T.$$

For an intermediate state we may then assume

$$\begin{aligned}\rho_S&=S\rho_1+(1-S)\rho_0\\&=\rho_d(1-S)+T\{\alpha_1 S+\alpha_0(1-S)\} \qquad \ldots.(8)\end{aligned}$$

This relation is compared with the experimental results of Kurnakow and Ageew [16] on Cu_3Au in figure 4. For the upper experimental curve the alloy

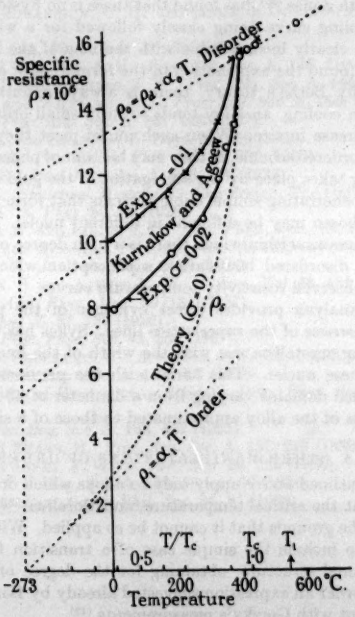


Figure 4. Cu_3Au. (Reproduced by courtesy of Prof. W. L. Bragg.)

was cooled in 3 hours and for the lower in 12–14 hours. The form of these curves suggests that if the alloy were annealed very slowly the experimental

curve might coincide with the theoretical, but at the same time an explanation of the considerable influence of the cooling rate is required.

The explanation was at first thought to lie in the fact that the interchange of A and B atoms, being a process involving an activation energy W, could not take place instantaneously and for any temperature there must be a finite time of relaxation

$$\tau = Ae^{W/kT}. \qquad \ldots.(9)$$

If the alloy is cooled too quickly it lags behind the equilibrium state and becomes "frozen" into a state of disorder. This effect may sometimes be apparent, but the case of Cu_3Au, for example, appears not so simple. Cu and Au atoms interchange quite readily at 250° C. while $T_c = 391°$ C., at which temperature, according to (9), atomic interchange would be so rapid that, however fast the alloy were quenched, ordering would occur. Further, Sykes, working with Evans [24] and with Jones [25], has found that there is no hysteresis on reheating the alloy, the cooling curve being closely followed for a wide range of temperature. This is clearly incompatible with the idea of the freezing-in of disorder, and Sykes found the explanation in the formation of *antiphase nuclei*.

As is shown by Bethe's theory, there is always a finite degree of local order σ. Thus, on cooling, an alloy tends to form small ordered nuclei which will gradually increase in size. When such nuclei meet they may coalesce to form a consistent ordered scheme, or they may be "out of phase". For example, in Cu_3Au ordering takes place by the segregation of the gold atoms on any one of the four inter-penetrating simple cubic systems that form the whole lattice, and the system chosen may be different in different nuclei. Thus the alloy is broken up into domains within which there is a high degree of order but which are separated by disordered boundaries, a conception which accounts satisfactorily for the observed resistivity-temperature curves.

Again, X-ray analysis provides direct evidence of the presence of these nuclei in the diffuseness of the superlattice lines. Sykes has used the Scherrer formula correlating crystallite size with the width of the lines to estimate the average size of these nuclei. This varies with the previous heat treatment, and he has obtained domains varying from a diameter of 40 A. to one so large that the properties of the alloy approximated to those of a single crystal.

§ 5. OTHER MANIFESTATIONS OF ORDER

The theories outlined above apply only to alloys which do not change their lattice structure at the critical temperature, and Borelius [22] has criticized the Bethe theory on the grounds that it cannot be so applied. Wilson [27], however, has extended it to include the simple case of a transition from face-centred cubic to a tetragonal structure, obtaining for the degree of tetragonality as a function of the order an expression suggested already by Borelius [3] as giving excellent agreement with Gorsky's measurements [12].

The method of X-ray analysis has been extensively applied to study the effect of plastic deformation on order, Linde [17] having thus demonstrated the presence of an ordered state in Cu–Pt which is undetected by other tests.

Other references to this work will be found elsewhere [20]. In the researches on the dependence of ordering upon composition the X-ray studies of the Fe–Al system may be mentioned. Bradley and Jay [4] thus deduced the distribution of Al atoms for concentrations up to 50 atomic per cent Al. Ferromagnetic properties are considerably affected by ordering, as has been shown by several workers, among whom Bradley and Rodgers have already been mentioned.

The general discussion of the above mentioned topics, as well as of many other interesting phenomena, such as, for example, the hysteresis effects that are in many cases associated with ordering, must be considered as outside the scope of this article. For the further developments the reader is again referred to the reviews [20], [22] and the many original papers mentioned therein.

REFERENCES

(1) Bethe, H.A. *Proc. Roy. Soc.* A, **150**, 552 (1935).
(2) Borelius, C., Johansson, C. H. and Linde, J. O. *Ann. Phys., Lpz.*, **86**, 291 (1928).
(3) Borelius, G. *Ann. Phys., Lpz.*, **20**, 57 (1934).
(4) Bradley, A. J. and Jay, A. H. *Proc. Roy. Soc.* A, **136**, 210 (1932).
(5) Bradley, A. J. and Rodgers, J. W. *Proc. Roy. Soc.* A, **144**, 340 (1934).
(6) Bragg, W. L. and Williams, E. J. *Proc. Roy. Soc.* A, **145**, 699 (1934).
(7) Chang, T. S. *Proc. Roy. Soc.* A, **161**, 546 (1937).
(8) Dehlinger, U. and Graf, L. *Z. Phys.* **64**, 359 (1930).
(9) Easthope, E. E. *Proc. Camb. Phil. Soc.* **33**, 502 (1937).
(10) Eisenschitz, R. *Proc. Roy. Soc.* A, **168**, 546 (1938).
(11) Fowler, R. H. *Statistical Mechanics.* (Cambridge: The University Press, 1936.)
(12) Gorsky, W. S. *Z. Phys.* **50**, 64 (1928).
(13) Johansson, C. H. and Linde, J. O. *Ann. Phys., Lpz.*, **78**, 439 (1925).
(14) Jones, F. W. and Sykes, C. *Proc. Roy. Soc.* A, **161**, 440 (1937).
(15) Kirkwood, J. G. *J. Chem. Phys.* **6**, 70 (1938).
(16) Kurnakow, N. S. and Agéew, N. W. *J. Inst. Met.* **46**, 481 (1931).
(17) Linde, J. O. *Ann. Phys., Lpz.*, **30**, 131 (1937).
(18) Moser, H. *Phys. Z.* **37**, 737 (1936).
(19) Mott, N. F. *Proc. Phys. Soc.* **49**, 258 (1937).
(20) Nix, F. C. and Shockley, W. *Rev. Mod. Phys.* **10**, 1 (1938).
(21) Peierls, R. *Proc. Roy. Soc.* A, **154**, 207 (1936).
(22) "The Conduction of Electricity in Solids." *Proc. Phys. Soc.* **49** (Extra Part), 77 (1937).
(23) Sykes, C. *Proc. Roy. Soc.* A, **148**, 422 (1935).
(24) Sykes, C. and Evans, H. *Proc. Roy. Soc.* A, **157**, 213 (1936).
(25) Sykes, C. and Jones, F. W. *J. Inst. Met.* **58**, 255 (1936).
(26) Wang, J. S. *Proc. Roy. Soc.* A, **168**, 56 and 68 (1938).
(27) Wilson, A. H. *Proc. Camb. Phil. Soc.* **34**, 81 (1938).

MEASUREMENTS WITH ULTRA-SHORT ELECTRIC WAVES ON WIRES

By L. HARTSHORN, D.Sc.
National Physical Laboratory

§ 1. INTRODUCTION

METHODS of measurement involving the use of electric waves of a wave-length of the order of a metre were very widely used forty or fifty years ago for investigations into the properties of liquid dielectrics, and scores of papers were published at that time giving data concerning the refractive index and absorption index for most of the available liquids for wave-lengths of this order. At that time the only known generators were of the spark type, so that all the measurements were made with damped waves, which is unfortunate, for, as Malsch [1] has recently shown, the use of damped waves is apt to introduce considerable errors, especially when the liquid under investigation shows large absorption. The development of generators of undamped waves of all wave-lengths from a few millimetres upwards has recently made it possible to make considerable improvements in the accuracy of measurements of this kind. The new technique has been applied not only to measurements of dispersion and absorption in liquids, but also to measurements of the dielectric constant and conductivity of soil, to determinations of magnetic permeability at very high frequencies, and to general measurements of impedance at wave-lengths down to 50 cm., a region which is steadily increasing in technical importance. The present *Report* reviews recent developments in those methods of measurement in which waves travelling along wires or parallel conductors are employed.

§ 2. LAVILLE'S EXPERIMENTS

A detailed experimental study of the propagation of continuous waves along a pair of parallel wires forming a transmission line was made some years ago by G. Laville [2], and as his work has obviously inspired some of the more recent papers, it will be convenient to deal with it first. Like many other observers he made his observations on a portion of the twin conductor terminated at either end by a conducting bridge. An e.m.f. of frequency $\omega/2\pi$ introduced into the line at any point gives rise to progressive waves of the form

$$u = Ue^{-\alpha x} \sin(\omega t - \beta x),$$

where α is the attenuation constant, and $\beta = 2\pi/\lambda$ the wave-length constant of the line. The waves are reflected by the conducting bridges at the ends, and stationary waves may be established. If a constant e.m.f. is induced into one of the terminal bridges (the input bridge) and the other bridge is

moved along the line so as to vary x, the distance between the bridges, i.e., the effective length of the line, then in the ideal case in which the two bridges are perfect reflectors, the current I in the input bridge varies in accordance with the equation

$$I^2 = A\frac{\cosh 2\alpha x + \cos \beta x}{\cosh 2\alpha x - \cos \beta x},$$

which defines a curve which, as x increases, rises at regular intervals to a maximum value of I, forming, as it were, a series of resonance curves. The most important case for measurement purposes is that in which the damping is very small, so that we may put

$$\cosh 2\alpha x = 1 + 2\alpha^2 x^2$$

and obtain

$$i^2 = A\frac{1 + 2\alpha^2 x^2 + \cos 2\beta x}{(1 - \cos 2\beta x) + 2\alpha^2 x^2},$$

the curve for which is as shown in figure 1. As x increases, the successive maximum values diminish rapidly in magnitude. The position of the nth maximum is given by

$$x_n = n\left(\frac{\lambda}{2} - \frac{\alpha^2\lambda^2}{8\pi^2}\right).$$

Thus the maxima are equally spaced, and the distance between any successive two is approximately equal to half the wave-length, λ. This relation provides

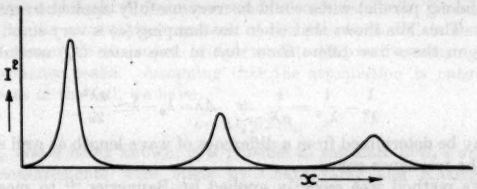


Figure 1.

the basis for the measurement of λ. The magnitude of the nth maximum is approximately given by $4A/\alpha^2\lambda^2 n^2$. Thus the diminution in the maxima as we pass along the line depends only on n, and not on α: all are equally reduced by an increase of α. It follows that α cannot be measured by observations of the maxima alone. It may, however, be determined from the sharpness of the resonance peaks. Laville shows that the equation for the nth peak, when its maximum point is taken as origin, may be written approximately

$$I^2 = A\Big/\left[\frac{n^2\alpha^2\lambda^2}{4} + \frac{4\pi^2x^2}{\lambda^2}\right],$$

from which it follows that the curve is symmetrical, and the values of x required to reduce I^2 to half its maximum value are $\pm n\alpha\lambda^2/4\pi$, i.e., the width of the nth peak at half its maximum height is $n\alpha\lambda^2/2\pi$, from which α may be determined.

Laville made measurements on a long line consisting of two parallel wires about 1 mm. diameter and 2 cm. apart, stretched between ebonite blocks, and he used wave-lengths of about 3 to 6 metres. The oscillator was inductively coupled to the line near one end, and at this end the two wires were connected by a fixed metal bridge in the middle of which was a thermo-element, connected to a sensitive galvanometer, the deflexion of which was proportional to I^2. Such an input bridge is obviously not a perfect reflector, but as it is stationary the departure from the ideal can be represented by a constant term in the equation. The moving bridge was made as perfect a reflector as possible. It consisted of a copper sheet with its plane perpendicular to the plane of the two wires, which passed through closely-fitting holes near the middle of the sheet. The size of sheet required was found by trial, successive additions being made until no further change in the resonance curve was produced. This bridge was provided with a micrometer adjustment, and settings near the resonance peaks could be made within a small fraction of a millimetre. For this arrangement the above equation is modified to

$$I^2=A\Big/\Big(\frac{n\alpha^2\lambda^2}{4}+n\alpha\lambda C+B+\frac{4\pi^2x^2}{\lambda^2}\Big),$$

where B and C are constants involving the properties of the line and the input bridge. Laville was able to verify this equation and to determine α, B, and C for various wires. He was also able to show that the equations of G. Mie [3] for waves along parallel wires could be very usefully applied to measurement purposes. Thus Mie shows that when the damping (α) is very small, the wave-length (λ) on the wires differs from that in free space (λ_0) according to the equation

$$\frac{1}{\lambda^2}-\frac{1}{\lambda_0^2}=\frac{\alpha}{\pi\lambda} \quad \text{or} \quad \Delta\lambda=\lambda_0-\lambda\simeq\frac{\alpha\lambda^2}{2\pi}.$$

Thus α may be determined from a difference of wave-length as well as from the width of the resonance curve.

Laville's method was recently applied by Bannerjee [4] to measurements of the "radiation resistance" of a line such as might be used for connecting an oscillator to an aerial system. The attenuation constant α of the line was obtained from a measurement of the width of the resonance peaks, and then the effective resistance was calculated by application of the usual formula $\alpha=\frac{1}{2}R\sqrt{(C/L)}$. The increase of radiation resistance produced by separating the wires or by bending them in various ways was measured in this manner.

§ 3. MEASUREMENTS ON SOIL

Measurements of the electrical properties of soil by means of stationary waves on wires immersed in the soil were first made by Smith-Rose and McPetrie [5]. Their measurements were limited to a determination of the ratio of the wave-lengths in air (λ_a) and in soil (λ_s) over a range of frequencies. If the wires are of negligible resistance we have

$$\lambda_a^2/\lambda_s^2=\tfrac{1}{2}[\epsilon+\sqrt{(\epsilon^2+4\sigma^2/f^2)}],$$

where ϵ is the dielectric constant of the soil, and σ is its effective conductivity

at the frequency f, so that limiting values for ϵ and σ can be estimated from the results. The measuring equipment was somewhat similar to Laville's, but it was found to be possible to avoid the inclusion of a detecting instrument in the input bridge. The sudden rise of current in this bridge at a resonance peak reacts on the oscillator and causes a sudden change in its anode current, and this change serves to indicate the position of the potential nodes as the movable bridge of copper sheet is moved along the wires. The distance between successive nodes gives $\lambda_a/2$. It is obviously impossible to move a bridge of copper sheet through the soil. The plan adopted was to pass the wires through a wooden box into which the soil could be put, and to place the box between the input bridge and the movable bridge so that the waves were reflected from the movable bridge after passing through the box. The box being empty, the first node on the far side of it was located. Soil was then gradually added to the box, thereby causing the node to travel up to and into it, owing to the contraction of wave-length in the soil. In this way several nodes could be observed to travel into the box as it was gradually filled, and by counting them, the distance between successive nodes in the box when full of soil could be determined. The front of the box was always set at a potential node so as to minimize disturbance due to reflexion at the surface of the soil.

The same technique for the determination of λ_a/λ_s was recently employed by Bannerjee and Joshi [6] for measurements on Indian soil. In addition they measured α, the attenuation constant of the line, by Laville's method, using a thermo-element in the input bridge as detector, and measuring the width of the resonance peaks. Assuming that the attenuation is entirely due to energy losses in the soil, we have

$$\sigma = \alpha\lambda_a^2 f/2\pi\lambda_s,$$

and λ_a, λ_s, and α being known, it is possible to calculate both ϵ and σ. Very similar measurements were made by Chakravarty and Khastgir [7] using Laville's technique, except that the movable bridge was of sufficiently small cross-section to permit of its being moved within the box of soil by means of a handle projecting through a slot cut in the top of the box.

It is only to be expected that the effective conductivity of soil will vary from sample to sample. Thus Smith-Rose [8] has reported values varying by a factor of 10^5 for surface soil from various parts of Great Britain, and it is therefore not very surprising that the values given by the Indian observers [6], [7] are of the order of a hundred times smaller than the corresponding ones given by Smith-Rose for Teddington. The differences in the effect of water content on the dielectric constant are, however, surprising. All the observations suggest that the dielectric constant of dry soil at a wave-length of a few metres is never very different from 4, but whereas in one case the presence of 20 per cent of water (the normal amount) increases this value to no more than 5 [7], in other cases the same quantity of water raised the dielectric constant to 15 or more [5], [6]. The difference is possibly related to particle size and adsorption.

§ 4. FERROMAGNETISM AT HIGH FREQUENCIES

Laville made comparative measurements of wave-length and attenuation constant α for waves on copper and iron wires. Both pairs of wires were about 1 mm. in diameter and 2 cm. apart, and the two lines were loosely coupled to the same oscillator and set at right angles so as to avoid interaction and reaction on the oscillator. The value of α for the copper wires at a wave-length of 7 m. was of the order of 1×10^{-5}, so that the wave-length was indistinguishable from that in vacuo, λ_0. For the iron wires the value of α was of the order of 20×10^{-5}, and it was therefore not only easily measured by observing the width of the resonance curve, but it was accompanied by a diminution of wave-length, $\Delta\lambda=\lambda_0-\lambda$, which was sufficiently large to permit the deduction of a value of α from this quantity also. Adopting Mie's [3] equation for the relation between α and the permeability μ, resistivity ρ, diameter d, and spacing D of the wires, viz.,

$$\alpha=(\mu\rho v/\lambda_0)^{\frac{1}{2}}/2vd\ln(2D/d),$$

Laville was able to deduce the permeability μ of the iron for high-frequency fields. For soft iron he found $\mu=74$, and the value was independent of frequency in the range $\lambda=3$ to 6·5 m., and apparently also up to 90 m. The magnetizing force was estimated to have the value 0·27 oersted. For nickel he obtained a value of $\mu=28{\cdot}5$.

More recently Lindman [9] has made further studies of this kind for much shorter wave-lengths. His technique was in many ways different from Laville's. The measurements were based on the diminution of wave-length only, and in order to make this large and easily measurable, very thin wires were used (diameter 0·02 mm.), and μ was deduced from Sommerfeld's formula for a single wire,

$$(\lambda_0-\lambda)/\lambda_0=3{\cdot}6\times10^{-8}\times(\rho\mu\lambda_0)^{\frac{1}{2}}/d.$$

He used a single pair of long parallel wires, which throughout most of their length were of copper and 1·3 mm. in diameter. Very thin wires of other metals were substituted for the thicker copper wire over a certain length of the line, and the length of stationary waves formed on these thin wires was compared with that on the thick copper wires, the latter value being indistinguishable from that in free space. The nodes and antinodes were detected by means of a tuned thermal detector (see below), which was moved parallel to the wires at a constant distance from them. The method obviously involves considerable difficulties, e.g., those due to reflexion at the points where the diameter of the wires changes, but the results obtained over the range $\lambda=16$ to 90 cm. are very interesting. For iron μ was found to have a constant value of 100 from $\lambda=90$ cm. to $\lambda=50$ cm., but at lower wave-lengths there was a rapid fall in permeability, a value of 10 being reached at $\lambda=16$ cm. For nickel [10] the permeability had the value 20 from $\lambda=90$ cm. to $\lambda=50$ cm., and then fell, reaching a value of about 2 at $\lambda=16$ cm. These observations are fairly consistent with those of Laville at longer wave-lengths, and also with a theoretical estimate by Becker [11] of the region in which ferromagnetic permeability might be expected to diminish considerably with rise of frequency.

§ 5. MEASUREMENTS OF ABSORPTION AND DISPERSION IN LIQUIDS

Methods involving the use of stationary waves on conductors have also been used for the measurement of absorption and dispersion in liquids such as water and the alcohols. A pair of parallel conductors is used, the input end being located in air, and the distant end with its terminal reflector immersed in the liquid, so that waves pass along the wires first through air and then through a layer of liquid terminated by a reflecting plate. The thickness of the layer of liquid through which the waves pass is easily controlled by moving the reflector along the conductors within the liquid. Waves incident on the first liquid surface are partly reflected and partly transmitted, and the transmitted portion suffers absorption in the liquid and reflection at the metal plate. The reflected wave travels back towards the input end, a portion reaching the detector, which is located in air on the input side of the liquid. As the thickness of liquid is varied, the deflexion of the detector passes through maxima and minima according as the reflected waves reinforce or oppose the incident waves. This method has been employed by Knerr [12] for measurements on water at wave-lengths of 20 to 6·5 cm. He shows that the expression for the response of the detector as a function of thickness of the liquid layer is somewhat complicated by the occurrence of multiple reflexion within the layer. However, by increasing the thickness until the absorption in crossing the liquid is so great that waves after crossing more than twice are of negligible intensity, it is possible to avoid undue complication. . In the simplest case the spacing of the maxima and minima on the curve connecting deflexion of detector and movement of reflecting plate is uniform and equal to one-quarter of the wave-length in the liquid. Since absorption increases with the thickness of liquid traversed, the maxima and minima diminish in value with increase of thickness, and the ratio of the amplitudes of the periodic term in the detector curve at successive maxima and minima serves to determine the absorption index of the liquid.

Malsch [13] has employed a somewhat different method. His apparatus is essentially similar, but the layer of liquid is so thick that any wave passing into it is completely absorbed and cannot affect the detector, which therefore responds only to the incident waves and those following from successive reflexions at the air-liquid surface and the input bridge. Vertical conductors are used, and the liquid reflecting-surface is moved along the wires by raising or lowering the level of the liquid. Mercury is first used as the liquid and the level adjusted until resonance is established in the usual way. The liquid under investigation is then substituted for mercury. The amplitude and phase of the reflected wave being now different, the liquid level must be displaced in order to restore resonance, and the magnitude of the displacement obviously serves to determine the change of phase on reflexion at the liquid surface. The reduction of amplitude on reflexion at the liquid surface affects the detector in much the same way as attenuation due to loss of energy in the line itself, and therefore diminishes the sharpness of resonance. The width of the resonance curve therefore serves to determine the amplitude of the reflexion coefficient. Then, having determined both amplitude and phase of the

reflexion coefficient, the refractive index and absorption index are calculated by application of Fresnel's formulae.

The shortest wave-length at which measurements have been made by these so-called wire waves appears to be 6·5 cm. Measurements at shorter wave-lengths have been made by using free waves and optical technique. Space does not permit of their discussion here, but it may be noted that diffraction is a serious difficulty in any attempt to apply optical technique to waves of such a length, and that wire-wave methods have the great advantage that the concentration of energy between the wires minimizes diffraction effects. In this connexion the work of G. C. Southworth [14] is very suggestive. It is shown that electric waves may be transmitted along dielectric cylinders or through hollow metal pipes in air, a pipe of any given diameter transmitting only wave-lengths less than a certain critical value, a pipe of 13 cm. diameter transmitting only waves of length shorter than 15 cm., and so on. Short sections of these "wave guides" resonate somewhat like the parallel wire systems already discussed, and since both radiation losses and resistance are very low, the resonance is very sharp. Waves may be generated by an oscillator valve mounted in such a section and transmitted along a pipe to a detector mounted in a similar section at the other end. The resonant cavity may be used as a wavemeter or as a means of measuring the properties of dielectrics placed within it. Few details are as yet available, but the scheme has attractive features.

§ 6. IMPEDANCE MEASUREMENTS

The measurement of the impedance of circuit components at very high frequencies by means of stationary waves on parallel conductors has recently been investigated by several workers. The methods employed vary considerably in detail, though they all depend on a comparatively few properties of the Lecher system. As R. A. Chipman [15] has pointed out, the parallel transmission line in virtue of its characteristic impedance (which may be calculated from its linear dimensions) forms an absolute standard of resistance, and observations on stationary waves virtually allow unknown impedances to be compared with this standard. Incidentally this illustrates G. A. Campbell's [16] contention that it would be possible to define the ohm in terms of the properties of free space.

In most measurements of this kind, the unknown impedance is connected across the parallel wires to form a bridge. A second bridge of negligible impedance, e.g. a large copper disc, is also provided, and resonance in the portion of the wires between the two bridges is observed. It is then found that the length of line between the two bridges when resonance occurs determines the unknown reactance, and therefore the capacitance or inductance, while the resistance or energy loss associated with it may be deduced either from the sharpness of the resonance curve obtained by varying the length of line, or alternatively from a measurement of the voltage across the unknown impedance at resonance.

Measurements of reactance by this method have recently been described

by E. Djakov [17], H. Kaufmann [18], R. King [19], and R. A. Chipman [15], among others. Their methods differ considerably in details; e.g., Djakov and King couple the generator inductively to the line at a point between the two bridges, while Kaufmann induces an e.m.f. into an additional short-circuiting bridge on an extension of the line, of length $\lambda/4$; and corresponding differences appear in the working equations, which tend to be less simple than could be wished. Djakov, Kaufmann and Chipman all deduce the resistive component of the unknown impedance from the width of the resonance curve. King uses a substitution method, depending on the ratio of the voltage at resonance across the unknown impedance to the voltage across a standard resistor which is substituted for it. Kaufmann also uses a "deflexion" method, in addition to the "resonance curve" method. He measures the ratio of the voltage across a standard resistor at resonance, to the corresponding voltage when the unknown impedance is connected in parallel with the standard. Provided the susceptances are very small, the resonance voltage is inversely proportional to the total conductance in the two cases, and therefore a very simple formula is obtained.

W. Hempel [20] has also described a deflexion method of impedance measurement, but his arrangement differed from the others previously mentioned

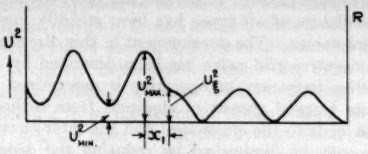


Figure 2.

in that the detector, instead of being connected to one of the bridges and stationary, was made capable of movement along the line, and his method depends on the ratios of deflexions at different points on the line, instead of at the same fixed point. The maxima and minima of voltage are noted as the detector traverses the line, which is terminated by an adjustable reflecting plate, and the voltage nodes and loops are thus located. The unknown impedance Z is then connected across the line at a voltage loop, resonance again established, and the maxima and minima of voltage at points along the line on either side of the unknown impedance measured. Provided the energy losses in the line itself are negligible, the voltage distribution is as shown in figure 2. On the portion of the line between Z and the terminal reflector R the voltage at the nodes is zero, and the value at the antinodes is U_ξ say. On the part of the line between the input-end and Z, the line voltage varies between the extreme values U_{max} and U_{min}, the latter only having the value zero in the case when Z is purely reactive. If Z_0 is the characteristic impedance of the line, it is shown that the magnitudes of Z and Z_0 are related by the equation

$$\left|\frac{Z}{Z_0}\right|^2 = \frac{U_\xi^2}{U_{max}^2 + U_{min}^2 - U_\xi^2} = b_1^2, \text{ say.}$$

Since the deflexion of the detector is usually proportional to U^2, this equation is convenient for the determination of the magnitude of Z. The reactance X associated with Z is determined by the distance x_1 (figure 2), which may also be measured, the appropriate relation being

$$X = \tfrac{1}{2} Z_0 . (1 - b^2) \tan 2\beta x_1 .$$

These methods all appear to have been successfully used at wave-lengths of less than a metre. The resonance-curve method appears to be capable of considerable accuracy. Kaufmann working at a wave-length of 50 cm. claims an accuracy of better than 1 per cent for resistances up to 1 000 Ω, and an accuracy of 6 per cent at 10 000 ohms, the maximum value for which the method is suitable. Higher resistances may be measured by the deflexion method.

§ 7. GENERATORS

It has been mentioned that progress in measurements with short waves has been greatly facilitated by the development of new generators of undamped waves. Several types are now available. The magnetron and oscillators of the Barkhausen-Kurtz type with positive-grid valves were described in volume I of these *Reports* (1934). In the last few years the limiting wave-length obtained with oscillators of all types has been steadily lowered by the construction of special valves. The development in this direction of the ordinary oscillator with a negative-grid valve has been described by A. L. Samuel [21]. The upper limiting frequency of operation is determined by the following factors: the finite time of transit of electrons from cathode to anode, the inductance of the leads to the electrodes, and the inter-electrode capacitance. The transit time may be diminished by reducing the distance between the electrodes, or increasing the anode voltage, but a limitation to these procedures is set by the maximum permissible emission per unit cathode surface, and the maximum anode dissipation. The inductance of the leads to the electrodes is best reduced by the use of straight leads of the shortest possible length, and the capacitances of the valves are reduced by keeping such leads far apart. A triode of ordinary size made on these lines forms a satisfactory oscillator for frequencies up to 700 Mc./s., while valves made of the smallest practicable dimensions (the " acorn " type) will oscillate at frequencies up to 1 000 Mc./s. The development of the magnetron as a short-wave oscillator of high power has been described by G. R. Kilgore [22]. Here again transit time is an important limitation, and it is reduced by the use of a small anode radius, a high anode voltage with special cooling arrangements, and a strong magnetic field. Further, in order that the oscillating circuit may be made as small as possible it is built inside the tube. O. Pfetscher and W. Puhlmann [23] have obtained an output of 80 watts at a wave-length of 19 cm. with a water-cooled magnetron of the negative-resistance type, while C. W. Rice [24] has obtained several watts at a wave-length of 5 cm. with an oscillator of the " transit time " type. When only very small amounts of power are required, it is possible to obtain still shorter wave-lengths. The possibilities have been thoroughly explored by

C. E. Cleeton [25] and N. H. Williams [26]. They used split-anode magnetrons, consisting of a straight wire cathode surrounded by a split cylindrical anode, the two halves of which were directly connected to a miniature Lecher wire system also enclosed within the tube. The available wave-length was diminished step by step by making tubes with smaller and smaller electrode-spacing, and increasing the anode voltage and the applied magnetic field. The shortest wave-length obtained was 0·64 cm., the anode radius then being only 0·019 mm., the length of the Lecher frame 0·38 cm., the anode voltage 1 200, and the magnetic field 24 000 oersted. Such a small generator naturally radiates very little power and, therefore, has few applications, but it is clear that measurements with undamped waves at wave-lengths down to, say, 2 cm. are now quite practicable.

§ 8. DETECTORS

The problems of the detection of very short waves, and of comparative measurements of their intensity, does not appear to present any special difficulty. Detectors of many kinds have been successfully used, including thermojunctions and bolometers as current indicators, and thermionic and crystal rectifiers as voltage indicators. For wave-lengths of the order of a few cm. the crystal rectifier appears to be the most convenient detector. Cleeton used an iron pyrites *vs.* phosphor bronze contact, while Southworth recommends silicon. At somewhat longer wave-lengths, several observers have used vacuum thermojunctions as current indicators by connecting them in the conducting bridge of the Lecher system. Hempel used the thermoelement as voltage detector by connecting it in the terminal bridge closing an auxiliary Lecher system of length $\lambda/4$, the open end of this system being placed near but not touching the main Lecher system. The detector system could thus be moved along the main system so that the maxima and minima of voltage along it could be measured. The whole arrangement may be regarded as a tuned thermal detector. Somewhat similar devices, but of very small linear dimensions, have been used as detectors of radiation of 4 cm. wave-length. D. Elle [27], for example, describes a bolometer for this purpose, consisting of a short length of Wollaston wire of diameter 2μ, its ends carrying extension pieces forming a small antenna, the overall length being adjusted so as to tune the detector to receive the required wave-length. The complete detector was mounted in a small bulb which could be evacuated. Djakov employed an " acorn " triode as thermionic voltmeter, while Kaufmann used a special diode of small linear dimensions. Although the actual calibration of the detector is not required, a knowledge of the law of its scale is required for most of the measurements. The thermal detectors may be assumed without doubt to follow a square law, but the law of the rectifiers cannot be taken for granted. It is interesting to note that Djakov found that his triode voltmeter of range 0·3 v. obeyed the square law at frequencies up to 500 Mc./s., and the same law was used by Kaufmann for his diode instrument. Knerr, working with free waves of wave-length 5 to 20 cm., found that his crystal rectifier (iron pyrites *vs.* tungsten) responded linearly to energy with an accuracy of

1 per cent, the law being the same as that reported by previous observers for low frequencies.

REFERENCES

(1) Malsch, J. *Ann. Phys., Lpz.*, **19**, 707 (1934).
(2) Laville, G. *Ann. Phys., Paris*, **2**, 328 (1924).
(3) Mie, G. *Ann. Phys., Lpz.*, **2**, 201 (1900).
(4) Bannerjee, S. S. *Phil. Mag.* **19**, 787 (1935).
(5) Smith-Rose, R. L. and McPetrie, J. S. *Proc. Phys. Soc.* **46**, 649 (1934).
(6) Bannerjee, S. S. and Joshi, R. D. *Phil. Mag.* **25**, 1025 (1938).
(7) Chakravarty, M. K. and Khastgir, S. R. *Phil. Mag.* **25**, 793 (1938).
(8) Smith-Rose, R. L. *Proc. Phys. Soc.* **47**, 923 (1935).
(9) Lindman, K. F. *Z. tech. Phys.* **19**, 158 (1938).
(10) Lindman, K. F. *Z. tech. Phys.* **19**, 323 (1938).
(11) Becker, R. *Z. tech. Phys.* **19**, 542 (1938).
(12) Knerr, H. W. *Phys. Rev.* **52**, 1054 (1937).
(13) Malsch. *Ann. Phys., Lpz.*, **20**, 33 (1934).
(14) Southworth, G. C. *J. Appl. Phys.* **8**, 661 (1937), and *Proc. Inst. Radio Engrs, N.Y.*, **25**, 807 (1937).
(15) Chipman, R. A. *Phys. Rev.* **53**, 672 (1938).
(16) Campbell, G. A. *Physics*, **3**, 230 (1932).
(17) Djakov, E. *Hochfrequenztech. u. Elektroakust.* **53**, 54 (1939).
(18) Kaufmann, H. *Hochfrequenztech. u. Elektroakust.* **53**, 61 (1939).
(19) King, R. *Proc. Inst. Radio Engrs, N.Y.*, **23**, 885 (1935).
(20) Hempel, W. *Elekt. Nachr.-Tech.* **14**, 33 (1937).
(21) Samuel, A. L. *J. Appl. Phys.* **8**, 677 (1937).
(22) Kilgore, G. R. *J. Appl. Phys.* **8**, 677 (1937).
(23) Pfetscher, O. and Puhlmann, W. *Hochfrequenztech. u. Elektroakust.* **47**, 105 (1936).
(24) Rice, C. W. *Gen. Elect. Rev.* **39**, 363 (1936).
(25) Cleeton, C. E. and Williams, N. H. *Phys. Rev.* **50**, 1091 (1936).
(26) Williams, N. H. *J. Appl. Phys.* **8**, 655 (1937).
(27) Elle, D. *Ann. Phys., Lpz.*, **30**, 354 (1937).

IMPEDANCE NETWORKS

By R. S. RIVLIN, B.A.
The Research Laboratories of The General Electric Company, Limited, Wembley, England

§1. INTRODUCTION

The subject of impedance networks has in the past thirty years attracted a great deal of attention from mathematicians, physicists, and engineers, as is evidenced by the many hundreds of papers that have been published dealing with various aspects of the subject. A variety of techniques has been applied, such as the operational calculus of Heaviside[1]; the application of matrix theory to network problems by Feldtkeller and Strecker [2], [3] among others; the use of continued fractions by Bartlett [4], Cauer [5], and Fry [6]; and the methods of synthesis applied to the design of wave filters by Cauer [7], [8], Bode and Dietzold [9], [10], and Gewertz [11].

The growth in the last few years of carrier telephone and telegraph systems and of the electrical "grid" system ensured that these techniques did not remain of academic interest only, but formed the basic methods by means of which physicists and engineers could solve practical problems.

The volume of published work on impedance networks is so great that it would be impossible to deal adequately with the whole field. We propose, therefore, to omit any discussion of transient phenomena and to limit ourselves to the advances which have been made in the theory and design of networks from the point of view of their steady-state behaviour. Even with this limitation of the field many further omissions, both intentional and unintentional, will be inevitable.

In his paper entitled "Electric Wave Filters", which was published in a previous volume of these *Reports*, N. F. Astbury [12] outlined the filter theory of Campbell and Zobel, and defined such terms as "propagation constant", "iterative impedance", "attenuation constant", "phase constant", and "image impedance".

An introduction to networks of the lattice or bridge types was also given in that *Report*, and networks of this type possessing the property of a constant, resistive iterative impedance were also mentioned; discussion of such networks figures largely in the present *Report*.

§2. GENERAL NETWORK THEORY

The study of networks has been approached, broadly, in two different ways, each possessing its own terminology. The method developed by the Americans, Campbell and Zobel, was followed by the German method, formulated by Feldtkeller [2], [3] and employed by Cauer and many others.

The latter network theory is unfamiliar to the majority of British and

American readers, but it is so important that it justifies more than a brief description. Actually a section in a recently published book by Guillemin [13] deals with the subject.

The general linear electrical network, or impedance network, consists of any combination of resistances, inductances, and condensers, which obeys a law analogous to Ohm's Law. Any sources of e.m.f. which the network may contain may be taken outside the network, and the two points at which the resultant e.m.f. is applied become terminals of the network.

We have, then, a passive network with two or more terminals and one or more meshes. Further, the network may be dissipative or non-dissipative according as it does or does not contain resistances. The former case is of interest in the design of attenuation-equalizing networks and in the study of the transmission range of wave filters.

In the most general case, we can write down, by using Kirchhoff's Laws, a set of simultaneous equations, linear in current and e.m.f., connecting the applied e.m.f.s $V_i (i=1, 2, \ldots n)$ and the currents $I_i (i=1, 2, \ldots n)$ in the n terminated meshes of the network (i.e. the n meshes which are accessible for the making of measurements).

Thus we have n equations,

$$\sum_{k=1}^{n} Z_{ik} I_k = V_i, \qquad \ldots.(1)$$

where Z_{ik} is an operator of the form

$$L_{ik}\frac{d}{dt} + R_{ik} + \frac{1}{C_{ik}}\int .dt,$$

for a branch containing an inductance L, resistance R, and capacity C in series.

We are mainly interested in the steady-state conditions which are obtained from the solution of the particular integrals of these differential equations, the complementary function giving rise to the transient terms.

In the four-terminal network, the terminals can be taken in pairs, so that one pair is considered as the input and the other as the output.

In this case equations (1) reduce to the form

$$\left.\begin{aligned} V_1 &= Z_{11} I_1 + Z_{12} I_2, \\ V_2 &= Z_{21} I_1 + Z_{22} I_2, \end{aligned}\right\}, \qquad \ldots.(2)$$

in which the Z coefficients have the dimensions of impedances. This is one form of the so-called network equations. Since we are considering only the steady-state behaviour, these equations are linear simultaneous equations for any single applied frequency.

Three other pairs of network equations which are mathematically identical with the equations (2) are

$$\left.\begin{aligned} V_1 &= A_{11} V_2 + A_{12} I_2, \\ I_1 &= A_{21} V_2 + A_{22} I_2, \end{aligned}\right\} \qquad \ldots.(3)$$

$$\left.\begin{aligned} I_1 &= Y_{11} V_1 + Y_{12} V_2, \\ I_2 &= Y_{21} V_1 + Y_{22} V_2, \end{aligned}\right\} \qquad \ldots.(4)$$

$$\left.\begin{aligned} V_1 &= D_{11} V_2 + D_{12} I_1, \\ I_2 &= D_{21} V_2 + D_{22} I_1, \end{aligned}\right\} \qquad \ldots.(5)$$

where the A, Y, and D coefficients can be determined in terms of the impedances of the arms of the network.

For the equations (3), the alternative notation

$$\left.\begin{aligned} V_1 &= A_1 V_2 + B_1 I_2, \\ I_1 &= C_1 V_2 + D_1 I_2 \end{aligned}\right\} \quad \ldots\ldots(6)$$

will be used.

Each of these pairs of equations is found to be most useful for one particular type of calculation ; for example, equations (2) are of most use when dealing with a system in which the input and output terminals of several four-terminal networks are connected separately in series (figure 1 *a*). Similarly, equations (3)

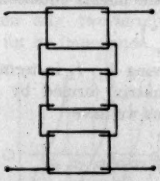


Figure 1 *a*. Three four-pole networks in series.

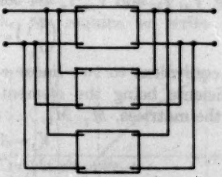


Figure 1 *c*. Three four-pole networks in parallel.

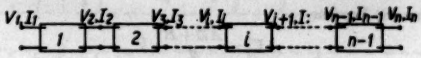


Figure 1 *b*. $(n-1)$ four-pole networks in tandem.

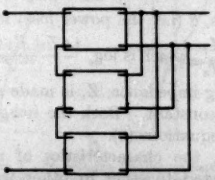


Figure 1 *d*. Three four-pole networks in series-parallel.

are employed in the case where several networks are connected in cascade or tandem (figure 1 *b*) ; equations (4) for the case in which the inputs of several networks and the outputs are connected separately in parallel (figure 1 *c*) ; and (5) for the case in which the input terminals of several networks are connected in series and the output terminals in parallel (figure 1 *d*).

Equations (6), for example, may be re-written in matrix form thus

$$\begin{pmatrix} V_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} A_1, & B_1 \\ C_1, & D_1 \end{pmatrix} \begin{pmatrix} V_2 \\ I_2 \end{pmatrix}.$$

If we have $(n-1)$ four-terminal networks connected in cascade as shown in figure 1 *b*, where V_i, I_i are the input potential difference and current respectively for the ith network and the output potential difference and current

respectively for the $(i-1)$th network, then we have $(n-1)$ equations connecting the input potential difference and current with the output potential difference and current for each of the $(n-1)$ networks. Thus, with a change of suffixes from those of equations (6), we have

$$\begin{pmatrix} V_i \\ I_i \end{pmatrix} = \begin{pmatrix} A_i, & B_i \\ C_i, & D_i \end{pmatrix} \begin{pmatrix} V_{i+1} \\ I_{i+1} \end{pmatrix}$$
$$= (M_i) \begin{pmatrix} V_{i+1} \\ I_{i+1} \end{pmatrix},$$

where i takes the values $1, 2, \ldots (n-1)$, and (M_i) denotes the matrix $\begin{pmatrix} A_i, & B_i \\ C_i, & D_i \end{pmatrix}$.

Hence, V_1, I_1, and V_n, I_n are connected by the matrix equation

$$\begin{pmatrix} V_1 \\ I_1 \end{pmatrix} = \prod_{i=1}^{n-1} (M_i) \begin{pmatrix} V_n \\ I_n \end{pmatrix}, \qquad \ldots.(7)$$

which is equivalent to two linear equations giving V_1, I_1 in terms of V_n, I_n, the coefficients being the elements of the matrix formed by multiplying together the matrices, M_1, M_2, $\ldots M_{n-1}$. Thus we have

$$\left.\begin{aligned} V_1 &= AV_n + BI_n, \\ I_1 &= CV_n + DI_n. \end{aligned}\right\} \qquad \ldots.(8)$$

Then, for any terminating impedance Z,

$$V_n = ZI_n,$$

and equations (8) become

$$V_1 = \frac{AZ+B}{Z} . V_n,$$

$$I_1 = (CZ+D) . I_n.$$

The transmission loss, θ (i.e. the power loss) for the termination Z is

$$\log_e \frac{V_1 I_1}{V_n I_n}, \text{ which is } \log_e \frac{(AZ+B)(CZ+D)}{Z}.$$

When the terminating impedance, Z, is made equal to the iterative impedance, θ is the transfer constant. Both the image and iterative impedances may be calculated from equations (8).

By similar methods, the characteristics of networks which consist of simpler networks connected together in other ways than cascade, may be found by starting with the pair of equations (2), (4) or (5).

The matrices for some very simple networks can be found by inspection and hence by application of the method outlined above, but based on equations (2) to (5). Hence, the matrices for certain standard types of section (e.g. lattice, T and Π) have been found and tabulated. By a further application of these methods, the matrices and hence the characteristics of networks which are combinations of these standard types of network may be found.

Further, the networks may contain controlled sources of e.m.f., such as thermionic valves working on linear portions of their characteristics, in addition to inductances, capacities, and resistances.

The conditions for the validity of the " matrix " method of solving network problems have been given by Baerwald [14].

§ 3. TWO-TERMINAL NETWORK

If in any network consisting of inductances, capacities, and resistances, we are concerned with the behaviour measured at two points of the network only, we may consider that we are dealing with a two-terminal network. The fundamental property governing its behaviour at these points is its (complex) impedance. This is defined in the following way :—if a potential difference represented by $Ee^{j\omega t}$ is applied between the two points, then the current flowing is represented by $Ee^{j\omega t}/Z$, where Z is the impedance of the two-terminal network.

In general, Z is a complex function of the frequency of the applied potential. The problems which arise in this connection are :—

(1) Given any two-terminal network, we require to write down the expression for its impedance.

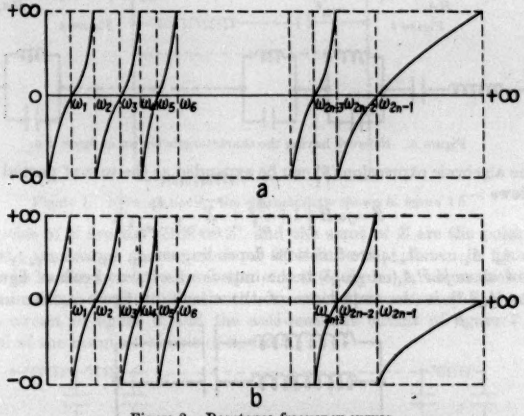


Figure 2. Reactance-frequency curves.
(a) for impedance Z; (b) for impedance Z', the reciprocal of Z.

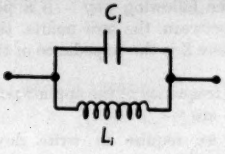
(2) Given a function of frequency (usually of the angular frequency, ω), to find a configuration of inductances, capacities, and resistances (and preferably the simplest configuration) which will have an impedance with this frequency characteristic.

The first problem is dealt with in elementary text-books by the use of the j notation. The second is, however, more difficult, and has been tackled in a number of ways.

Foster [15] proved the theorem that a two-terminal network consisting of pure inductances and capacities has, in many cases, the impedance

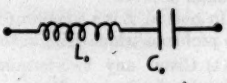
$$Z=\frac{j\omega H(\omega^2-\omega_1^2)(\omega^2-\omega_3^2)\ldots.(\omega^2-\omega_{2n-1}^2)}{\omega^2(\omega^2-\omega_2^2)(\omega^2-\omega_4^2)\ldots.(\omega^2-\omega_{2n-2}^2)}. \qquad \ldots.(9)$$

The impedance has zeros for angular frequencies $\omega_1, \omega_3, \ldots \omega_{2n-1}$ and poles for angular frequencies $0, \omega_2, \omega_4, \ldots \omega_{2n-2}$. The shape of the impedance characteristic is determined by these frequencies and the scale by H (figure 2 *a*). Conversely, Foster showed how any impedance characteric of this form can be realised, in practice, with the greatest economy of elements.



$$C_i = -\frac{1}{HA_i}, \quad L_i = -\frac{HA_i}{\omega_{2i}^2}.$$

Figure 3.



$$L_0 = H, \quad C_0 = -\frac{1}{HA_0}.$$

Figure 4.

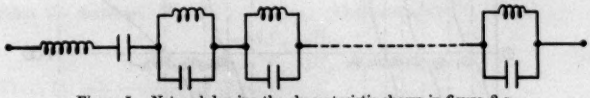


Figure 5. Network having the characteristic shown in figure 2 *a*.

The algebraic expression (9) can be expanded as the sum of partial fractions as follows :—

$$Z = j\omega H\left[1 + \frac{A_0}{\omega^2} + \sum_{i=1}^{n-1} \frac{A_i}{\omega^2 - \omega_{2i}^2}\right],$$

where $A_0, A_1, \ldots A_{n-1}$ are functions depending on $\omega_1, \omega_3, \ldots \omega_{2n-1}$ only.

Now since $j\omega HA_i(\omega^2 - \omega_{2i}^2)$ is the impedance of the circuit of figure 3 and $j\omega H[1 + A_0/\omega^2]$ is the impedance of the circuit of figure 4, the impedance

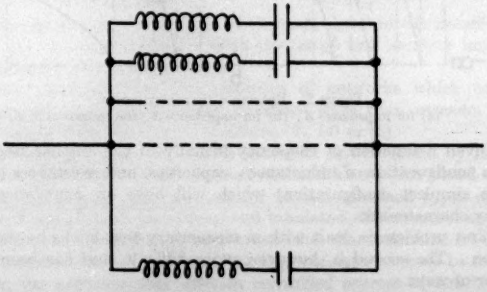


Figure 6. Alternative network having the characteristic shown in figure 2 *a*.

may thus be realised by the configuration of figure 5, which contains $2n$ components arranged as a resonant circuit in series with a number of anti-resonant circuits.

An alternative realisation, also consisting of components, of the impedance Z is shown in figure 6. This is obtained by expressing $1/Z$ as the sum of partial fractions.

The reciprocal impedance Z', of Z with respect to a resistance R, may be defined by

$$ZZ' = R^2.$$

If Z has the form (9), then Z' has the form

$$\frac{\omega(\omega^2-\omega_2^2)(\omega^2-\omega_4^2)\ldots.(\omega^2-\omega_{2n-1}^2)}{jH(\omega^2-\omega_1^2)(\omega^2-\omega_3^2)\ldots.(\omega^2-\omega_{2n-2}^2)}.$$

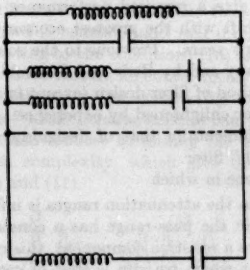


Figure 7. Network having the characteristic shown in figure 2 *b*.

The poles of Z are the zeros of Z', and the zeros of Z are the poles of Z', so that the impedance characteristic of Z' will be that shown in figure 2 *b*, and may be realised practically as the circuit of figure 7, where each of the resonant circuits is the reciprocal with respect to the resistance R of an anti-resonant circuit of figure 5, and the anti-resonant circuit of figure 7 is the reciprocal of the resonant circuit of figure 5.

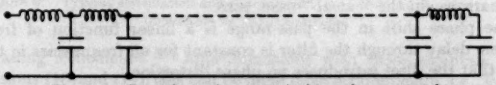


Figure 8. Cauer form of two-terminal impedance network.

Cauer [5] has shown that the impedance

$$Z = \frac{\alpha_{2n}\lambda^{2n}+\alpha_{2n-2}\lambda^{2n-2}+\ldots.+\alpha_2\lambda^2+\alpha_0}{\beta_{2n-1}\lambda^{2n-1}+\beta_{2n-3}\lambda^{2n-3}+\ldots.+\beta_3\lambda^3+\beta_1\lambda},$$

where $\lambda = j\omega$,

may be expanded as a continued fraction of the form

$$a_1\lambda + \cfrac{1}{b_1\lambda + \cfrac{1}{a_2\lambda + \cfrac{1}{\ddots + \cfrac{1}{a_n\lambda + \cfrac{1}{b_n\lambda}}}}},$$

and can therefore be realised as the ladder network of figure 8, which consists of n sections each containing two elements.

It can be seen that the Foster and Cauer realisations of any impedance

characteristic employ the same number of elements, and this is the minimum number of elements with which the impedance can be realised.

Both Foster [16] and Cauer [5] have extended their methods to the physical realisation of impedance : frequency characteristics by networks employing resistances as well as reactances.

§ 4. FILTER SYNTHESIS

We have seen that the characteristics of a given filter may be calculated by the matrix method from a knowledge of the elements.

The inverse problem of this, i.e., the problem of calculating the elements of the filter which will give a required performance as regards attenuation, impedance, and phase shift with the greatest economy, has attracted a good deal of attention in recent years. Previous to the solution of these problems by Cauer [7], [8] in Germany and by Bode and Dietzold [9], [10] and Gewertz [11] in America, the only method of filter design (except in the very simplest cases) was that of trial and error enlightened by experience.

The problem is fundamentally that of designing with a given degree of approximation an " ideal " filter.

An " ideal " filter is one in which

(1) The attenuation in the attenuation ranges is infinite.

(2) The attenuation in the pass-range has a constant zero value. If the elements of the filter have a resistive component, this cannot be achieved, and the " ideal " condition at which we aim is that of constant, low attenuation in the pass range.

(3) Since the filter is normally terminated in a resistance which is independent of frequency, this requires that the " ideal " filter should have a constant, purely resistive, characteristic impedance in the pass-range so that no reflexion losses dependent on frequency are introduced by the termination.

(4) The " transition ranges " of frequency in which the attenuation rises from its value in the pass-range to its infinite value in the attenuation ranges are very narrow—in the " ideal " case, zero.

(5) The phase shift in the pass-range is a linear function of frequency, i.e. the time delay through the filter is constant for all frequencies in the pass-range, so that the filter introduces no phase distortion.

The " ideal " filter would require for its realisation a network having an infinite number of elements and therefore, in practice, cannot be made. However, we can approximate to it as closely as we please by making our network sufficiently complex.

Cauer takes as his " ideal " filter, a filter having the characteristics (1) to (4) inclusive, and does not attempt to approximate to a linear phase-shift characteristic. Bode and Dietzold, however, define the " ideal " filter as having all five of the properties given above, and make the approximation to a linear phase-shift *vs.* frequency characteristic the prime consideration.

Cauer's Synthesis [7], [8]

In Cauer's synthesis, a lattice filter, whose line and lattice arms have impedances Z_1, Z_2 respectively, is considered.

The characteristic impedance, Z, of such a filter is given by $Z=\sqrt{Z_1 Z_2}$, and the attenuation in decibels A, by the formula

$$A=20\log_{10}\frac{1+\sqrt{Z_1/Z_2}}{1-\sqrt{Z_1/Z_2}}.$$

We see that for the " ideal " filter,

$$Z_1/Z_2=1 \text{ in the attenuation ranges} \quad \ldots.(10)$$

and $$Z_1 Z_2=R^2 \text{ (where } R \text{ is a constant pure resistance) in the pass-range.} \quad \ldots.(11)$$

We can thus approximate to the conditions that A should be infinite in the attenuation ranges and Z constant and real (i.e., purely resistive) in the pass-range, by approximating to the conditions (10) and (11).

Z_1/Z_2 and $Z_1 Z_2$ must have certain algebraic forms containing parameters, the number of which determines the complexity of the network, and the closeness with which the " ideal " characteristics can be approached. Mathematically, the problem then reduces to finding the actual parameter values, in a network of a given complexity, which give the closest approximation to the conditions of (10) and (11).

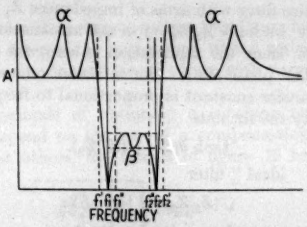


Figure 9. Typical attenuation and characteristic. Impedance characteristics obtainable by Cauer synthesis.

By making Tchebichev [17] approximations with these algebraic forms to the conditions (10) and (11), this can be done most economically, i.e., for a given number of parameters and hence for a given complexity of the lattice network, compliance with the conditions (10) and (11) is most closely approached.

By multiplication and division of the expressions for Z_1/Z_2 and $Z_1 Z_2$ so obtained, we can obtain expressions for Z_1 and Z_2.

From these formulæ for Z_1 and Z_2, which are actually of the form of (9), the physical configuration of the arms of the filter can be found by Foster's Reactance Theorem and are of the type shown in figure 5.

The type of attenuation characteristic obtained by this synthesis for a band-pass filter with cut-off frequencies f_1 and f_2 is shown in figure 9. The curve α is the attenuation characteristic outside the frequency range $f_1 f_2$. The curve β represents (on a different scale) the characteristic impedance in the frequency range $f_1 f_2$.

For the attenuation characteristic, we see that outside the range $f_1' f_2'$, the attenuation lies above a certain value A' and that each of the dips descends to this value. By increasing the complexity of the arms of the filter, the transition intervals $f_1'f_1$ and f_2f_2' may be made as narrow as we please and the attenuation level A' may be made as high as we please. Actually, as we increase the complexity (by increasing the number of anti-resonant circuits in the series of figure 5), the number of dips in the attenuation range increases.

From the curve β, figure 9, we see that inside the frequency range $f_1'' f_2''$, the characteristic impedance oscillates between two limiting values which may, by increasing the complexity of the filter, be made as close together as we please. Further, the intervals $f_1 f_1''$ and $f_2 f_2''$ may be made as narrow as we please.

The distinctive feature of the Tchebichev method of approximation is shown by the fact that all the dips between successive peaks of attenuation fall to the same level A' and that the characteristic impedance oscillates in such a way that all its maxima have the same value and all its minima have the same value.

Bode and Dietzold's Synthesis [9], [10]

Here again a lattice filter with arms of impedances Z_1 and Z_2 is considered. For an ideal filter, we have $Z_1/Z_2=1$ in the attenuation ranges.

In the pass-range, since the attenuation is zero, the transfer constant, θ, is proportional to the phase-shift. For a linear phase-shift *vs.* frequency characteristic the transfer constant is proportional to frequency.

Now, θ is given by the formula

$$\tanh \theta/2=\sqrt{(Z_1/Z_2)}.$$

Therefore, in the " ideal " filter

$$\sqrt{(Z_1/Z_2)}=j \tan \pi f/2\alpha$$

throughout the pass-range, where α is a constant.

Then $\sqrt{(Z_1/Z_2)}$ may be expressed as an infinite product, for

$$j \tan \frac{\pi f}{2\alpha} = \frac{j\pi f(1-f^2/2^2\alpha^2)(1-f^2/4^2\alpha^2)\ldots.}{2\alpha(1-f^2/\alpha^2)(1-f^2/3\alpha^2)\ldots.} \quad \ldots.(12)$$

This infinite product may be represented with any desired degree of accuracy by a finite product which for a band-pass filter with cut-off frequencies at f_a, f_b is, for example,

$$K\sqrt{(1-f^2/f_a^2)(1-f^2/f_2^2)\ldots.(1-f^2/f_{2n}^2)}\sqrt{(1-f^2/f_b^2)}$$
$$(1-f^2/f_1^2)(1-f^2/f_3^2)\ldots.(1-f^2/f_{2n+1}^2)$$

The number of the critical frequencies $f_1, \ldots . f_{2n+1}$ determines the closeness with which the network may be made to approximate to the condition of linear phase-shift characteristic.

The characteristic impedance may be made to approximate to a constant resistance and the physical configuration may be determined in the manner previously described in connexion with Cauer's synthesis.

Filters designed by the method of Bode and Dietzold have, unlike those of Cauer, an attenuation which rises steadily as we proceed from a cut-off frequency into the attenuation range. As the condition of a linear phase-shift characteristic in the transmission range is more and more closely approached, the attenuation rises more and more rapidly in the attenuation range.

However, in so far as it is desired purely to obtain high attenuation in the attenuation range, this method is not as economical as that of Cauer, and the synthesis of Bode and Dietzold has its value primarily in the design of filters with a linear phase-shift characteristic in the transmission range.

Dietzold [18] and Norton [19] have given methods of approximating with a lattice filter to linear phase-shift characteristic, not only in the transmission range, but also in the adjacent regions of the attenuation ranges, and not only for termination of the filter in its characteristic impedance, but for any terminating resistance.

For practical purposes, these two methods are of greater utility than that of Bode and Dietzold, for filters are usually terminated in a resistance and not in their characteristic impedance.

§ 5. NETWORKS WITH CONSTANT INPUT IMPEDANCE

Filters

It has been shown in the previous section how Cauer has solved the problem of designing a single filter of the lattice type with a resistive input impedance which remains as nearly constant as we please with varying frequency, in the transmission range.

A number of methods of obtaining a constant purely resistive input impedance which depend on the use of a compensating network have been suggested. Amongst others, the circuit of figure 10 has been proposed [20].

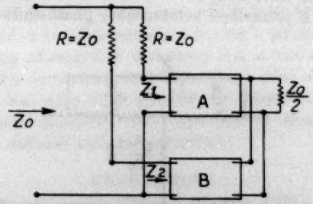


Figure 10. Method of obtaining constant input impedance using compensating network. (Siemens and Halske A.G. method.)

Here the filter A has an input impedance Z, and the input impedance Z_2 of the compensating filter B is made resistance-reciprocal with respect to A, so that it is given by

$$Z_1 Z_2 = R^2.$$

The input impedance of the composite network is then R. The two filters A and B have identical transmission properties and may be constructed, for example, as **T** and Π networks respectively.

In practice, it is often desired to operate a number of filters in parallel (figure 11) so that the input impedance to the whole group of filters is a constant resistance R. Norton [21] has shown that this may be achieved by making the input impedance of filter A, say, of the form $R[1+F(\omega)]$, subject to certain limitations on the form of the function $F(\omega)$, and the input impedance of the

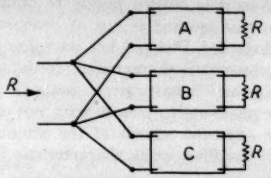


Figure 11. Method of obtaining constant input impedance using compensating network. (Norton's method.)

filters B, C, etc., operating in parallel of the form $R[1+1/F(\omega)]$. If necessary, a compensating network may be added in parallel with the filter group to make it possible to obtain this input impedance.

Darlington [22] has also proposed an interesting method of obtaining constant resistive input impedance.

Distortion-correcting networks

Fundamentally, a distortion-correcting network is a device for correcting the distortion produced in a signal by some apparatus which introduces an attenuation or phase-shift which varies with frequency. The problem of the design of such correcting networks then resolves itself into the design of networks having a prescribed attenuation, phase-shift or transfer-constant characteristic.

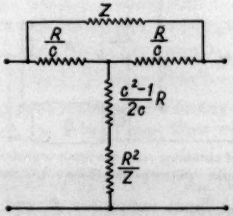


Figure 12. Zobel-Stevenson equaliser.

The design of such distortion-correcting networks having a constant purely-resistive input impedance has been discussed fully by Zobel [23].

Among many other types of distortion-correcting networks, Zobel describes a bridged-T network of the form shown in figure 12.

The transfer constant, θ, of such a network is given by the formula

$$e^{\theta}=\frac{1+(c+1)Z/2R}{1+(c-1)Z/2R}, \qquad \ldots.(12)$$

and the network is physically realisable only if the arbitrary constant c is greater than unity.

More recently, Feldtkeller [24] has shown that the network of figure 13 has a transfer constant, θ, given by the formula

$$e^{\theta}=\frac{1+\left(\frac{1}{c}+1\right)\frac{Z}{2R}}{1+\left(\frac{1}{c}-1\right)\frac{Z}{2R}}, \qquad \ldots.(13)$$

if the transformer windings are equal and wound in opposite directions.

Here again the network is physically realisable if c is greater than unity.

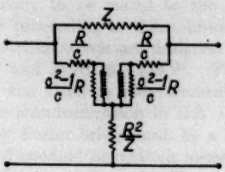


Figure 13. Feldtkeller's modification of Zobel-Stevenson equaliser.

We see, by comparing equations (12) and (13), that the network of figure 13 enables us to realise practically transfer constants which could be obtained theoretically with the network of figure 12 only by taking c less than unity.

Bode [25] has discussed the requirements which have to be met by, and the practical realisation of, a variable equaliser having a single varying resistance, suitable for correcting attenuation distortion (i.e. a variation of attenuation with frequency) which is temperature-dependent, over a given temperature range. Hitherto, it has only been possible to construct variable equalisers by employing variable inductive and capacitative elements of which the construction is often difficult and uneconomical.

REFERENCES

(1) Heaviside, O. *Electromagnetic Theory*. (Reissue, 1922.)
(2) Strecker, F. and Feldtkeller, R. *Elekt. Nachr.-Tech.* **6**, 93 (1929).
(3) Feldtkeller, R. *Einführung in die Vierpoltheorie der elektrischen Nachrichtentechnik* (1937).
(4) Bartlett, A. C. *Phil. Mag.* **49**, 1072 (1925).
(5) Cauer, W. *Arch. Elektrotech.* **17**, 355 (1927).
(6) Fry, T. C. *Bull. Amer. Math. Soc.* **35**, 463 (1929).
(7) Cauer, W. *Math. Z.* **38**, 1 (1933).
(8) Cauer, W. *Siebschaltungen* (1931).
(9) Bode, H. W. *J. Math. Phys. M.I.T.* **13**, 275 (1934).
(10) Bode, H. W. and Dietzold, R. L. *Bell Syst. Tech. J.* **14**, 215 (1935).

(11) Gewertz, C. M. *Network Synthesis* (1933).
(12) Astbury, N. F. *Rep. Progr. Phys.* **5**, 334 (1938).
(13) Guillemin, E. A. *Communication Networks* (1935), vol. II, chap. IV.
(14) Baerwald, H. G. *Elektr. Nachr.-Tech.*, **9**, 31 (1932).
(15) Foster, R. M. *Bell Syst. Tech. J.* **3**, 259 (1924).
(16) Foster, R. M. *Bell Syst. Tech. J.* **3**, 651 (1924).
(17) Tchebichev, P. L. *Mem. Acad. Sci. St. Petersb.* **9**, 201 (1859).
(18) Dietzold, R. L. U.S. Pat. Spec. No. 2,054,794.
(19) Norton, E. L. U.S. Pat. Spec. No. 2,070,677.
(20) Siemens and Halske, A. G. Brit. Pat. Spec. No. 462,984.
(21) Norton, E. L. *Bell Syst. Tech. J.* **16**, 178 (1937).
(22) Darlington, S. U.S. Pat. Spec. No. 2,115,138.
(23) Zobel, O. J. *Bell Syst. Tech. J.* **7**, 438 (1928).
(24) Feldtkeller, R. *Wiss. Veröff. Siemens-Konz.* **14** (3, 64) (1935).
(25) Bode, H. W. *Bell Syst. Tech. J.* **17**, 229 (1938).

THE MEASUREMENT OF CAPACITANCE

By N. F. ASTBURY, M.A., F.Inst.P.
National Physical Laboratory

§ 1. INTRODUCTION: FUNDAMENTAL METHODS

In the *Report* for 1934 a brief account was given of methods used for the so-called "absolute" measurement of capacitance. For the most part the work there described is of a highly specialized nature, representing as it does the methods by which the unit of capacitance is realized and maintained in the various national standardizing laboratories. In the present *Report* we shall be chiefly concerned with methods of capacitance measurement which are more likely to be useful to the general experimenter and research worker. Before proceeding to these, however, it is necessary briefly to refer to the new capacitance test-set designed at the National Physical Laboratory by Astbury and Ford in 1938 [1]. This set, while incidentally providing a method for the absolute measurement of capacitance, provides the basis for capacitance standardization in this country. The basic circuit is derived from the Carey Foster bridge and, by a combination of two bridge circuits, most of the small residual corrections necessary in the earlier versions of the method are completely eliminated. The normal range of the bridge, which is intended for frequencies below a few thousand cycles per second, is from 1000 $\mu\mu$F. to 20 μF., and over that range it is possible to obtain an accuracy of two or three parts in 10^5 on capacitance and 0·00003 radian on loss angle. These quantities are determined in terms of mutual inductance and resistance, both of which quantities are in turn directly related to the primary standard of mutual inductance of the Laboratory.

§ 2. CLASSIFICATION OF METHODS

A rough classification of general methods is worth attempting, although one is conscious of the extreme difficulty of adding recommendations as to choice of method. The problem may present itself in many forms—such, for example, as in essential measurements in the physico-chemical study of dielectrics, in the measurement of minute mechanical displacements, in the recording of thickness, humidity, etc., of non-conducting materials,—and the responsibility for a choice of method is best left to the experimenter who is conversant with all the aspects of the problem.

Broadly speaking, there appear to be the following possibilities for capacitance measurement:—

(i) *Bridge methods.* These methods probably represent the widest range: bridges can be produced to handle capacitances from a fraction of a micro-microfarad to many microfarads and to cover a frequency range from power to radio-frequencies.

(ii) *Resonance methods.* These are suitable for capacitances up to values

of the order of 1000 $\mu\mu$F. and for frequencies greater than about 50 kc./s. Both bridge and resonance methods, if they are to be properly exploited, call for some degree of specialized technique and are full of pitfalls for the tyro.

(iii) *Oscillator and valve methods.* These methods generally make use of valve circuits executing continuous or relaxation oscillations and may, by the inclusion of deflexional instruments, remove a lot of responsibility from the observer. They can be made to cover a very wide range of capacitance and frequency, although they are less precise in general than either bridge or resonance methods.

(iv) *Impedance, voltage and current methods.* These methods again may employ deflexional instruments and include both precision and "rough and ready" methods.

(v) *D.C. and ballistic methods.* These are clearly limited in scope, but may occasionally be useful and informative.

§ 3. BRIDGE METHODS

Very comprehensive surveys of bridge methods are given by Hague [2] and by Campbell and Childs [3] in their books. Versions of the Carey Foster circuit have been given by Campbell [4] and by A. W. Smith [5], both of whom have produced direct-reading arrangements. Watson [6] has shown that the Campbell-Heaviside bridge for comparing mutual and self inductance lends itself to the measurement of small capacitances. The capacitance under test is associated with a high resistance and one branch of the bridge in delta connexion, rather in the manner of the classical Anderson network. All these, of course, are exclusively audio-frequency bridges. The now-popular Schering bridge has been shown to be suitable for radio-frequency work [7], although Mau [8] considers it inferior at such frequencies to the four-condenser bridge, in which power-factor adjustments are made with a series resistance in one arm. Kind [9] also prefers this type of bridge but includes a series resistance in every arm. Series combinations in bridge arms may, however, lead to somewhat obscure and complicated effects [10], and on the whole they are best avoided wherever possible.

Bünger [11] has handled the four-condenser bridge in what appears to be a novel and effective manner. He supplies the bridge through a capacitive coupling arrangement from a quartz-controlled oscillator (480 m.) and builds up about 500 volts across it by tuning with a shunt inductance. Variation in the capacitive coupling supplies the quadrature (power factor) adjustment, the main (capacitance) adjustment being made with a differential condenser across two arms. A two-stage screen-grid amplifier, feeding into a rectifier and reflecting galvanometer (for photographic recording), serves as detector. The amplifier has a gain of 120 db. and is sensitive to 1 μv. Changes of capacitance of one part in a hundred million can be recorded to about 10 per cent. The bridge arms are of the order of 1000 $\mu\mu$F., so that the actual capacitance changes observed in Bünger's work were about 10^{-5} $\mu\mu$F. Such changes were not produced mechanically: amongst other things Bünger was investigating the relationship between dielectric constant and pressure for air.

§ 4. RESONANCE METHODS

The phenomenon of resonance in circuits consisting of capacitance and self-inductance furnishes a very attractive basis for capacitance measurements, but exacting practical requirements may have to be met before the best use can be made of this simple idea. Perhaps one of the best of recent studies of the subject is that due to Hartshorn and Ward [12]. In their dielectric test-set, which can be used up to 100 megacycles per second, the main variable condenser forming part of the resonating circuit is operated by a precision micrometer head, the whole very robust arrangement being built on to a valve voltmeter with self-contained reflecting galvanometer for detecting resonance. A second micrometer operating a plunger in one of the main electrodes provides supplementary capacitance adjustments to 0·001 $\mu\mu$F. This is used to measure the width of the resonance curve, from which loss factors can be deduced. The authors give elsewhere [13] some typical general measurements. A similar instrument has been described by Dunford and Goodall [14]. This, of course, is precision equipment and may not commend itself for general work. The papers should nevertheless be studied by anyone wishing to emulate the work on a less ambitious scale.

Often the use of the resonance curve at its peak, where the gradient is small, may be very disappointing. An arrangement due to Oliphant [15] circumvents this difficulty. A closed circuit of capacitance, inductance and resistance is loosely coupled to an oscillator and a valve voltmeter is connected across the condenser. This can be designated as circuit (1). A similar circuit, (2), is loosely coupled to (1) and provision is made for connecting the capacitance under test in shunt with the condenser in this circuit. Circuit (1) is set just off resonance : as circuit (2) is tuned through resonance the effective reactance of (1) varies sharply according to the known properties of coupled circuits. The curve connecting effective reactance of (1) with capacitance in (2) is shaped roughly like a capital N, the diagonal of which is practically straight. The middle point of this diagonal corresponds to the normal reactance of (1). The diagonal represents the working range of the equipment in which small changes in capacitance in (2) produce large changes of reactance, and hence of voltage, in (1). The slope of the diagonal, which determines the range and sensibility of the method, can be controlled by the resistance in circuit (2) and is large when this resistance is low. Oliphant has measured capacitance changes as small as 0·01 $\mu\mu$F. by this method and has shown that it can be made continuously recording by the use of an oscillograph.

Fromy [16] has described the direct use of a resonance method in which the test circuit is capacitively coupled to the oscillator, while Rutelli [17] has explored the possibilities arising from certain relationships shown to exist between specific changes in the three parameters of a resonant circuit. These relationships may take the form of parabolic or hyperbolic curves.

A method due to Rohde and Leonhardt [18] should perhaps be included here, although it is hardly a true resonance method. The current from a high-frequency oscillator is split into two anti-phased components which pass respectively through two branches, one of which contains the condenser under

test and the other a variable standard. The latter is adjusted to give zero resultant volts at the remote junction of the branches, where the resultant voltage is applied to a tuned circuit. The balance condition is judged from the lack of reaction on this circuit. The method was designed for measurements on inter-electrode capacitances of valves and covers a range from 0·001 $\mu\mu$F. to 10 $\mu\mu$F.

§ 5. OSCILLATOR METHODS

A search of the literature reveals a surprisingly large number of variations of the idea of using oscillating circuits for capacitance measurements. Roughly, the schemes fall into two groups—those in which oscillators (continuous or relaxation) are made to vary in frequency or output by the application of the unknown capacitance, and those in which the application of the unknown capacitance produces a change in the direct currents (anode currents) of the system.

The continuous oscillator methods may employ a fixed oscillator and a second oscillator, the frequency of which is altered by the insertion of the capacitance under test. Consonance between the oscillators, as indicated by the absence of "beats", is established before and after insertion of the unknown capacitance by adjustment of the calibrated tuning condenser of the second oscillator. From the scale readings of this instrument the value of the unknown is obtained by difference. Examples of this method in which a quartz crystal controls the fixed oscillator are given by Leonard [19], by Stranathan [20] and by Fritsch [21]. In an arrangement due to Lister [22] simple oscillators are employed and a visual indicator is provided which allows the oscillators to be brought not into absolute synchronism but into a condition in which their frequencies differ by a small predetermined amount of the order of 5 cycles per second. Sathe and Rangachari [23] describe the use of a single oscillating valve coupled to an anode-bend detector valve.

Circuits due to Fricke [24] and to Spencer Smith [25] provide examples of the amplitude variation method. The former describes a 5-megacycle oscillator very sensitive to changes in anode capacitance and suitable for measuring small mechanical displacements as capacitance changes: the latter used a dynatron oscillator with valve voltmeter or cathode-ray oscillograph for continuous recording. The range of these methods is about 10 $\mu\mu$F.

A very good example of the use of anode current change as a capacitance variation indicator is provided by Schmidt's study of a quartz oscillator [26]. As the capacitance in the tuned-anode circuit is varied, the direct anode-current falls very suddenly on the inception of oscillations: over a short range the direct current then falls linearly as the alternating component rises exponentially, dropping suddenly again as the tuned circuit comes into resonance with the crystal. Schmidt used a 3-megacycle oscillator, and the working range, represented by the interval between the two sudden changes in the anode current, was 10 $\mu\mu$F., and changes of 0·01 $\mu\mu$F. could be detected. Niemeyer [27] and Marketu [28], following Löwy [29], have used a simple tuned-anode oscillator for the anode-current variation method. Prytherch [30], in a dilatometer for

metallurgical work, used a series-fed Hartley oscillator in which mechanical movements were used to alter the grid capacitance. Anode current changes were recorded on a string galvanometer and displacements as small as 10^{-6} inch could be detected.

Valve circuits employing relaxation oscillations have been devised by Sesta [31] and by Guillien [32]. The former arranges two triodes so that the grid circuit of the first is transformer-coupled to the plate circuit of the second, the plate of the first being coupled to the grid of the second through a shunt combination of capacitance and resistance. This circuit executes relaxation oscillations the period of which is determined by the product of the capacitance and resistance in the coupling unit. The oscillations are slow enough to be conveniently timed by a stop-watch. In Guillien's relaxation oscillator a conventional tuned-grid valve circuit is modified by the insertion of a shunt combination of resistance and capacitance in series with the tuned circuit. In both cases the period of the relaxation oscillations is proportional to the capacitance over a very wide range.

The relaxation oscillations of the familiar condenser-resistance-neon-tube circuit form the basis of Durepaire's " phonic capacity meter " [33]. It is not, however, sufficient to carry out the measurements by substitution, because of the profound effect of condenser losses on the frequency of the oscillations. Durepaire shows how to circumvent this difficulty by putting the capacitance under test in series with a condenser of specially high insulation-resistance.

A " direct-reading " capacity meter due to Vecchiachi [34] uses two triodes the grids of which are supplied with antiphased a.c., so that the valves become alternately conducting and non-conducting every other half-cycle. The condenser under test is charged through one valve and discharged through the other, current being drawn from a high-tension battery. This current is directly proportional to the product of capacitance, frequency and battery voltage. This method and the preceding three are, in general, unsuitable for capacitances below a few thousand $\mu\mu$F.

§ 6. IMPEDANCE METHODS

The methods treated in this section are those which depend on some sort of measurement or indication of voltage across, or current through, the capacitance under test. A precision method of this type has been described by Astbury and Jones [35]. The capacitance to be measured is put in series with a calibrated variable condenser, and a capacitance potential-divider or attenuator is connected across the combination, forming a four-condenser bridge across which a constant high-frequency voltage is applied. A valve voltmeter (uncalibrated) is connected in turn across the variable condenser and across the output of the attenuator, the latter being adjusted so that the voltmeter reading remains unchanged during the transfer. This process is repeated for a few readings of the variable condenser so that a correct allowance may be made for the input capacitance of the voltmeter, and the unknown capacitance is deduced in terms of the readings of the variable condenser and the reduction ratio of the attenuator. The method can be used for capacitances

as low as 0·001 $\mu\mu$F. A somewhat similar method for use at power frequencies has been described by Pistoia [36].

In an arrangement described by Brown [37] gold-leaf electroscopes are used as indicating voltmeters. The unknown capacitance is put in series with a standard, and about 1000 v. (r.m.s.) is applied from a transformer connected to the 50~mains. An electroscope is connected across each condenser and intersecting images of the two gold leaves are projected on to a screen. The locus of the point of intersection under stated conditions of potential drop across the two condensers is a family of slightly curved lines. The "equivolt" lines can be traced on the screen and a kind of balance condition established. The method is naturally limited to those condensers which can successfully withstand the high voltage required for adequate sensitivity. Bruce and Wilcken [38], using the series arrangement of unknown and standard, compare the voltages with the aid of a cathode-ray oscillograph. The voltages across the two condensers are applied to the X and Y plates and a linear trace at fixed slope indicates equality. The method is particularly good for leaky condensers and is suitable for values from 1 000 $\mu\mu$F. to 1 μF.

Beck [39] in his capacity meter compares the currents through an unknown and a standard condenser. The currents are balanced against each other by a rectifier bridge, a d.c. meter serving as indicator. An ingenious use of transformers extends the range from 5 $\mu\mu$F. to 2 μF. without multiplicity of standards.

§ 7. D.C. METHODS

All d.c. methods of measuring capacitance are hallowed by tradition and one can do little more than report improvements rather than novelties. Astbury and Ford [40] have improved the ballistic method of comparing capacitance with mutual inductance and resistance, and Wright and Graham [41] have shown that a post-office box and a ballistic galvanometer of known constants can be used for capacitance measurement. Barton [42] has described a neat variation of the charge-sharing method in which the test condenser is alternately connected to a charged variable air condenser and discharged. A Wilson tilting electroscope is used to indicate the potential drop on the air condenser, which is adjusted after every charge-sharing operation so as to maintain its potential at a constant level. The test capacitance is $1/n$th of the total capacitance-change after n sharings. The author shows how to correct for leakage, and gives figures to show that the method is suitable for capacitances from 5 to 10 000 $\mu\mu$F.

§ 8. CONCLUSION

It will be evident that practically all the methods described above depend on the use of a calibrated standard. While it is outside the scope of the present *Report* to discuss generally the question of standards, a few points can be made to supplement the foregoing descriptions. Precision variable air-condensers of the moving-vane type ranging from 10 to 5000 $\mu\mu$F. are now made by most good instrument firms. For smaller capacitances the micrometer plunger condenser has found a lot of favour. Examples of this type, which

can be read to about 0·001 $\mu\mu$F., have been described by Ward and Pratt [43], Clay [44], Gill [45] and Bartlett [46]. Stranathan [47] boldly solves the small condenser problem by removing most of the plates from a worm-driven vane condenser, although one doubts whether mechanically this is likely to be comparable with the micrometer type. It may be worth pointing out here that condensers intended for high-frequency work should be free from appreciable self-inductance, and papers by Ward [48], by Field and Sinclair [49] and by Holzner [50] should be consulted for practical discussions of this problem.

Finally, there is an important distinction between bridge methods and most others which is not always fully appreciated. A condenser may consist effectively of three electrodes (one of which may be the earth or any earth-connected neighbouring body) only two of which provide the relevant capacitance. Bridge circuits can generally be adapted, by the use of artificial earth points, to measure such capacitances: the other methods, with the exception of that due to Rohde and Leonhardt [18], do not possess this property.

REFERENCES

(1) Astbury and Ford. *Proc. Phys. Soc.* **51**, 37 (1939).
(2) Hague. *A.C. Bridge Methods.* (Pitman, London, 1938.)
(3) Campbell and Childs. *The Measurement of Inductance, Capacitance and Frequency.* (Macmillan, London, 1935.)
(4) Campbell. *Proc. Phys. Soc.* **43**, 564 (1931).
(5) Smith. *Rev. Sci. Instrum.* **4**, 280 (1933).
(6) Watson. *Canad. J. Res.* (A), **13**, 5 (1935).
(7) Dye and Jones. *J. Instn. Elect. Engrs*, **72**, 169 (1933).
(8) Mau. *Arch. Elektrotech.* **31**, 473 (1937).
(9) Kind. *Elek. u. Masch.* **52**, 292 (1934).
(10) Astbury. *Phil. Mag.* **24**, 507 (1938).
(11) Bünger. *Z. Phys.* **91**, 679 (1934).
(12) Hartshorn and Ward. *J. Instn. Elect. Engrs*, **79**, 597 (1936).
(13) Hartshorn and Ward. *J. Sci. Instrum.* **14**, 132 (1937).
(14) Dunford and Goodall. *Wireless Engr.* **15** 662 (1938).
(15) Oliphant. *J. Sci. Instrum.* **14**, 173 (1937).
(16) Fromy. *Rev. Gen. Elect.* **39**, 611 (1936).
(17) Rutelli. *Ricerca Sci.* **2**, 364 (1938).
(18) Rohde and Leonhardt. *Arch. Tech. Messen.* **92**, V. 3532–4 (1939).
(19) Leonard. *Electronics*, **11**, 18 (1938).
(20) Stranathan. *Rev. Sci. Instrum.* **5**, 334 (1934).
(21) Fritsch. *Arch. Tech. Messen.* **92**, V. 3531–1 (1939).
(22) Lister. *Wireless Engr.* **11**, 425 (1934).
(23) Sathe and Rangachari. *Wireless Engr.* **8**, 543 (1931).
(24) Fricke. *Hochfrequenztech. u. Elecktroakust.* **43**, 149 (1934).
(25) Spencer Smith. *J. Sci. Instrum.* **12**, 316 (1935).
(26) Schmidt. *Hochfrequenztech. u. Elecktroakust.* **41**, 96 (1933).
(27) Niemeyer. *Phys. Z.* **31**, 451 (1930).
(28) Marketu. *Phys. Z.* **35**, 629 (1934).
(29) Löwy. *Phys. Z.* **26**, 649 (1925).
(30) Prytherch. *J. Sci. Instrum.* **9**, 128 (1932).
(31) Sesta. *Nuovo Cim.* **8**, 114 (1931).
(32) Guillien. *C.R. Acad. Sci., Paris*, **204**, 668 (1937).
(33) Durepaire. *Rev. Gen. Elect.* **34**, 39 (1933).
(34) Vecchiachi. *Alta Frequenza*, **1**, 202 (1932); *Proc. Inst. Radio. Engrs, N.Y.*, **19**, 659 (1931).

(35) Astbury and Jones. *J. Sci. Instrum.* **13**, 407 (1936).
(36) Pistoia. *Alta Frequenza*, **7**, 245 (1938).
(37) Brown. *Proc. Phys. Soc.* **44**, 220 (1932).
(38) Bruce and Wilcken. *Beama J.* **43**, 7 (1938).
(39) Beck. *Proc. Phys. Soc.* **45**, 765 (1933).
(40) Astbury and Ford. *J. Sci. Instrum.* **15**, 94 (1938).
(41) Wright and Graham. *J. Sci. Instrum.* **12**, 220 (1935).
(42) Barton. *Rev. Sci. Instrum.* **3**, 123 (1932).
(43) Ward and Pratt. *J. Sci. Instrum.* **16**, 192 (1939).
(44) Clay. *Physica*, **3**, 757 (1936).
(45) Gill. *J. Instn. Elect. Engrs*, **69**, 521 (1931).
(46) Bartlett. *J. Sci. Instrum.* **8**, 260 (1932).
(47) Stranathan. *Rev. Sci. Instrum.* **5**, 315 (1934).
(48) Ward. *J. Sci. Instrum.* **13**, 251 (1936).
(49) Field and Sinclair. *Proc. Inst. Radio. Engrs, N.Y.*, **24**, 255 (1936).
(50) Holzner. *Alta Frequenza*, **7**, 582 (1938).

DIELECTRIC BREAKDOWN IN SOLIDS

By H. FRÖHLICH

H. H. Wills Physical Laboratory, University of Bristol

OWING to its great technical importance, the dielectric breakdown in solids has for many years been a subject of experimental and theoretical investigations. Nevertheless, only in recent years has it been possible to come to a closer understanding of this phenomenon. It is the aim of this article to give an account of these recent developments.

One of the most important results of recent research has been the experimental proof of the existence of an *intrinsic electric strength*. This means that at a given temperature a maximum breakdown strength exists for each dielectric substance which is a constant of this substance, and which is obtained under ideal conditions (homogeneous field, uniform material without weak spots, etc.). Therefore it should be, and has been, possible to calculate this intrinsic electric strength from simple physical constants of the material. In this *Report* we shall deal mainly with the intrinsic electric strength. There exists also a quite different type of breakdown, the so-called breakdown through *thermal instability*. This type of breakdown is of importance only in special conditions such as high temperature. We shall give only a very short account of the thermal breakdown, since several reports on it have been published already [30], [26], [14]. In addition, there are the various complex forms of breakdown which occur in industrial insulation; but these, although of great engineering importance, usually reduce finally to one of the two fundamental types considered. The description of industrial breakdowns would be lengthy and is related to the particular properties of complicated substances. Since, also, there is an extensive engineering literature on the subject, it will not be treated here.

I. THE INTRINSIC ELECTRIC STRENGTH

§ 1. VON HIPPEL'S EXPERIMENTS

Several years ago von Hippel[16] published an account of a series of experiments on electrical breakdown, which were of very great importance for the understanding of this phenomenon and for the development of a theory of it. Von Hippel chose for his experiments * single crystals of alkali halides, and thus materials with very simple structure. He was able to show that the breakdown strength is almost the same for different specimens and does not depend on their source or method of preparation. He thus established certain values for the breakdown strength of alkali halides which we shall discuss in § 6. Moreover, he established the following facts:

* A description of his experimental technique was given in an earlier *Report* by Hartshorn [14].

(i) The breakdown strength of mixed crystals is always higher than that of their components. As an example we show in figure 1 the breakdown strength of KCl—RbCl, a series which forms a continuous range of solid solutions.

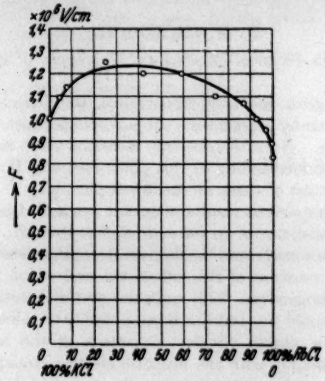


Figure 1. The breakdown strength of the mixed crystal KCl—RbCl according to von Hippel [16].

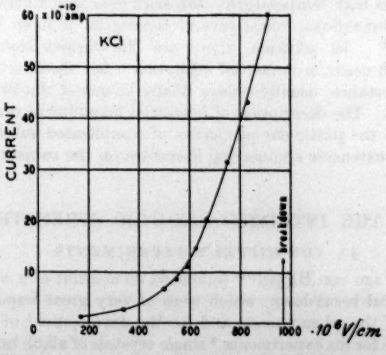


Figure 2. Electrical current in KCl in strong fields, according to von Hippel [17].

(ii) The breakdown occurs preferentially along certain crystallographic axes. For alkali halides the preferred axis is the 110.

(iii) He studied the electrical currents occurring at field strengths somewhat below the breakdown strength. These currents rise rapidly with increasing field strength (cf. figure 2). Nevertheless they are very weak even just below

the breakdown strength and, according to von Hippel, the current density increases almost discontinuously at the breakdown field.

When von Hippel's experiments were carried out no satisfactory theory was available. Various theories had been proposed but had failed to account for all the facts. The mechanical theory, for instance, assumes that the breakdown is due to mechanical rupture of the crystal, caused by the forces which the electrical field exerts on the ions. According to this theory the electrical strength, like the mechanical strength, should depend greatly on cracks or other crystal imperfections, which is in contradiction to von Hippel's experiments. On the other hand, Joffé has assumed * that the breakdown is due to an ionization of the ions by the moving ions, carrying the current. This theory does not allow, however, for the short time in which the breakdown occurs, which according to Rogowsky is of the order of 10^{-8} sec., neither does it seem plausible in the light of present theories of ionic conduction that the ions could acquire sufficient energy.

There remained the hypothesis, which was suggested by von Hippel and others, that the breakdown is a phenomenon due to electrons. This suggestion stimulated the present author to study theoretically the movement of electrons in dielectrics, and this led to a quantitative theory of electrical breakdown [8]. In the following we shall first give an outline of this theory and then discuss the results and some further experiments, carried out at the Laboratory of the British Electrical and Allied Industries Research Association (E.R.A.).

§ 2. THE MOVEMENT OF ELECTRONS THROUGH CRYSTALS

Before we turn to a study of the condition for breakdown, we shall discuss the general features of the movement of electrons through crystals. Let us assume an ideal crystal with all atoms or ions at rest at their respective lattice points. The behaviour of electrons in such crystals was first studied theoretically by F. Bloch [2] in connexion with the electron theory of metals.† Bloch's main result was that the energy spectrum of electrons in a crystal consists of bands of allowed energy levels which are separated by forbidden zones. According to the Pauli principle each energy level can be occupied by no more than two electrons (with anti-parallel spins). If such an energy band is completely occupied, the total current of its electrons is zero. Consequently the structure of an insulator must be such that its energy bands are fully occupied up to a highest band, and that the width of the next forbidden zone is so large that under normal conditions no electrons will be thermally excited into the first empty band. We shall call the levels in the empty energy bands, the *conduction levels* of the insulator. If electrons are brought into the conduction level of an insulator (say, for instance, by absorption of light) the crystal loses its insulating properties and an electrical current is set up in an external electrical field, proportional to the number of electrons in the conduction level.

It follows from theoretical considerations that the electrons in the conduction level behave in a way very similar to free electrons, whose kinetic

* Cf. the discussions in Semenoff and Walther's book [26].

† Cf. books on the Electron Theory of Metals [7], [20].

energy may be considered as zero at the bottom of the conduction band and which have a certain effective mass m^* instead of the electronic mass m. The exact value of the ratio m^*/m has not yet been calculated, owing to mathematical difficulties, but there are reasons for believing that m^*/m is approximately equal to unity (cf. Tibbs [27]).

So far, we have seen that an electron in the conduction band of an idealized crystal (i.e. with all atoms fixed at their lattice points) behaves similarly to a free electron in vacuum, i.e. it could be accelerated without any friction by an external electrical field. In order to account for the finite mobility of the electrons, we have, according to F. Bloch, to consider the deviations from the strict periodicity of the lattice. These are, in a pure crystal, mainly due to the thermal vibrations of the atoms or ions. These thermal vibrations give rise to a scattering of the electrons (as described in more detail in § 3) and hence to a finite mobility ω. ω is defined as the average drift velocity of the electron per unit field, i.e.

$$\bar{v}=\omega F, \qquad \ldots.(1)$$

where $\bar{v}$ is the drift velocity and F the field strength. $\bar{v}$ is of course zero without an external field, whatever the kinetic energy $\frac{1}{2}m^*v^2$ may be, since at each collision the velocity v changes its direction (cf. figure 4 *a*). Instead of the mobility, we may introduce the time of relaxation τ defined by

$$\tau=m^*\omega/e, \qquad \ldots.(2)$$

or the mean free path l determined by

$$l=\tau v. \qquad \ldots.(3)$$

In an external field each electron makes a contribution ev/V (V=volume of the crystal) to the total current density, which, according to (1) and (2), is given by

$$\frac{e\bar{v}}{V}=\frac{e^2\tau F}{m^*V}.$$

Thus, according to electrodynamics, the field transfers the energy

$$A=e^2\tau F^2/m^* \qquad \ldots.(4)$$

to the electron per second. A is, of course, the energy transfer in the average over several collisions of the electron. Between two collisions, the energy exchange between field and electron is proportional to F, and not to F^2, and may be positive or negative according to the direction of the velocity v of the electron relative to the field.

In order to calculate the time of relaxation τ, we have to consider the interaction between an electron and the lattice vibrations. Such a calculation is similar to those in the theory of metallic resistance first developed by F. Bloch. As in this theory, it is rather difficult for non-polar lattices to obtain quantitative theoretical values for τ. This is due to the fact that the forces acting upon an electron, which are due to the thermal vibrations of the atoms, are of a rather complicated form. This is, however, not the case in ionic crystals, where the thermal vibrations of the ions give rise to an electrical polarization of which the interaction with an electron is well known. Thus, in order to calculate τ for ionic crystals, we shall first study their thermal vibrations.

§3. LATTICE VIBRATIONS AND MEAN FREE PATH IN IONIC CRYSTALS

Consider a lattice of the NaCl type. Let u_+, u_- denote the displacements of the ions of the two signs in the neighbourhood of a given point r and let

$$u(r)=u_+-u_-.$$

Then, representing the ions approximately by point charges $\pm e$, the average electrical polarization P near r is given by

$$P(r)=eu(r)/2a^3,$$

where $2a^3$ is the volume of the unit cell, i.e. a is the distance between neighbouring ions. Consequently any displacement of the ions of a lattice can be expressed in terms of a polarization $P(r)$. Now each function $P(r)$ may be developed into a series of plane waves (Fourier series), which we shall call polarization waves. The shortest polarization waves have wave-lengths of the order of the lattice distance a. Still shorter wave-lengths would not have any meaning, owing to the atomic structure. There exist two kinds of polarization waves: transverse waves, for which div $P=0$, and longitudinal ones, for which div $P\neq 0$.

In the absence of external fields, the amplitude of a polarization wave varies harmonically with the time. Its frequency ν will in general be different for longitudinal and transverse waves and will also depend on the wave number w ($w=2\pi/$wave-length). The frequency of long transverse waves is known experimentally from infra-red spectroscopy (residual frequencies). The frequencies of longitudinal waves, which we shall need later on, are not known experimentally. It follows from theoretical considerations [11] that the frequency of short longitudinal waves is roughly (but not exactly) equal to the residual frequency ν_l. Long longitudinal waves, however, have a much higher frequency ν_l given by

$$\nu_l^2=(\epsilon-\epsilon_0+1)\nu_t^2, \qquad \ldots.(5)$$

where ϵ is the dielectric constant for static fields and ϵ_0 the contribution to the dielectric constant from the polarizability of the ions, so that $\sqrt{\epsilon_0}$ is the refractive index in the near infra-red.

According to quantum mechanics, the energy levels of each polarization wave with frequency ν are given by

$$(n+\tfrac{1}{2})h\nu, \quad (n=0, 1, 2, \ldots).$$

In thermal equilibrium, the average value $\bar{n}$ of n depends on the temperature T according to Planck's law

$$\bar{n}=\frac{1}{e^{h\nu/kT}-1}. \qquad \ldots.(6)$$

The interaction energy $W(r)$ between an electron at the position r and a polarization wave with wave-number w (polarization P_w) is determined by electrodynamics according to the differential equation

$$\nabla^2 W(r)=4\pi e \operatorname{div} P_w(r).$$

Therefore, only longitudinal polarization waves interact with an electron.

Now the scattering of an electron by a polarization wave is pictured by quantum mechanics in the following way: An electron may, due to the interaction W, either absorb or emit a polarization wave. The probability for such a process is proportional to n for absorption and proportional to $(1+n)$ for emission. Energy and momentum have to be conserved for each such process. Consequently, if K is the wave-number of the electron, i.e. if $\hbar = h/2\pi$,

$$E = \hbar^2 K^2/2m^*, \qquad \ldots.(7)$$

its kinetic energy before the transition, and K' and E' the respective quantities after the transition, we find

$$E' = E \pm \hbar\nu \qquad \ldots.(8)$$

and

$$K' = K \pm w. \qquad \ldots.(9)$$

Consider a fast electron with $E \simeq 5$ ev., i.e. if $m/m^* \sim 1$, $K \simeq 10^8$ cm^{-1}. Since $\hbar\nu$ is of the order of 1/10 ev., the collision is nearly elastic, i.e. $K' \simeq K$. It then follows from (9) that the scattering angle ϑ, i.e. the angle between K and K', is given by

$$\sin \vartheta/2 = w/2K. \qquad \ldots.(10)$$

Consequently, long polarization waves (small w) lead to a small scattering angle, but the scattering angle through collision with the shortest waves ($w \simeq 2\pi/a \simeq 10^8$ cm^{-1}) is large. Scattering, of course, can (cf. (10)) only occur if $w < 2K$. The mean free path l for such electrons is found to be [3]

$$l = l_0 \left(1 + \frac{2}{e^{h\nu/kT} - 1}\right)^{-1}, \qquad \ldots.(11)$$

where l_0 is the mean free path at $T = 0$:

$$l_0 = \frac{16}{2^{2/3}\pi^2} \frac{Ma^5\nu E^2}{e^4\hbar}, \quad \text{if } E > E_0. \qquad \ldots.(12)$$

Here

$$E_0 = \frac{2^{2/3}\pi^2}{8} \frac{\hbar^2}{m^* a^2}, \qquad \ldots.(13)$$

and M is the reduced mass of the ions whose masses are M_+ and M_-, i.e.

$$\frac{1}{M} = \frac{1}{M_+} + \frac{1}{M_-}.$$

The main contribution to l is due to the scattering by very short polarization waves. Therefore, ν is the frequency of short longitudinal waves, which is approximately equal to the residual frequency. Using this, an expression for l_0 may be derived without using the mass and charge of the ions, but using instead the dielectric constant ϵ and the refractive index in the near infra-red, $\sqrt{\epsilon_0}$:

$$l_0 = \frac{8}{2^{2/3}\pi^3} \frac{a^2 E^2}{(\epsilon - \epsilon_0) e^2 \hbar\nu}, \quad E > E_0. \qquad \ldots.(14)$$

This expression can easily be generalized for ionic crystals of a more complicated structure, which have more than one residual frequency.

According to equation (11), the mean free path increases with decreasing temperature, but reaches a finite value at $T = 0$, although no polarization

quanta are then present ($\bar{n}=0$). This remaining scattering is due to the zero-point vibrations, i.e. to the fact that the lowest energy state of an oscillator with frequency ν is not zero but $\frac{1}{2}h\nu$.

We shall later on need the mean free path for fast electrons only. For slower electrons no method has yet been found which permits a satisfactory calculation of the mean free path. For very slow electrons, i.e. $E<h\nu$, the method used for fast electrons can again be applied. But now only those collisions are possible where the electrons *absorb* an $h\nu$—an emission being impossible because $E<h\nu$. This causes some alterations in the derivation of the mean free path. The result is [11]

$$l=\frac{3(\epsilon-\epsilon_0+1)}{\epsilon-\epsilon_n}a_0\left(\frac{E}{h\nu_l}\right)^{1/2}(e^{h\nu_l/kT}-1), \quad E<h\nu_l, \qquad \ldots.(15)$$

where
$$a_0=\frac{\hbar^2}{me^2}=0{\cdot}54\times10^{-8}\text{ cm.}$$

is the Bohr radius and ν_l is given by equation (5). This theoretical value of the mean free path is in fair agreement with experimental values obtained from photoconductivity. This is of importance for us since the formula for fast electrons which we shall need for the calculation of the breakdown strength is derived by the same theoretical method.

As mentioned above, a fast electron may either absorb or emit a quantum $h\nu$ when it is scattered, the former process being proportional to $\bar{n}$, the latter to $(1+\bar{n})$. Apart from this factor, the probabilities for emission or absorption are the same. Consequently, the energy B transferred per second from an electron to the lattice is proportional to $(1+\bar{n})-\bar{n}$, i.e. independent of the temperature T, since $\bar{n}$ is the only quantity depending on T. B is found to be

$$B=\frac{h\nu}{\tau_0}\frac{E}{E_0}\log\gamma, \quad E>E_0, \qquad \ldots.(16)$$

where
$$\gamma=4(EE_0)^{1/2}/h\nu, \qquad \ldots.(17)$$

Here τ_0 is the time of relaxation at $T=0$, i.e. according to (3),

$$\tau_0=l_0/v,$$

where l_0 is given by equation (14). Thus the rate of loss of energy is slightly larger than $h\nu$ per τ_0 seconds. B is not exactly equal to $h\nu/\tau_0$, since τ_0 is not the exact time between two collisions, but also contains the mean scattering angle.

§ 4. THE CONDITION FOR BREAKDOWN [8], [9]

In the previous sections (2 and 3) we have seen how the movement of an electron in a lattice can be described by a mean free path l or a time of relaxation $\tau=l/v$. We found that in an external electrical field F an electric current is set up.

The conception of the mean free path, as calculated in § 3, only has a meaning if the field F is not too strong, namely if

$$eFl\ll E, \qquad \ldots.(18)$$

i.e. if the energy gain or loss of an electron between two collisions is small compared with its kinetic energy. This energy transfer is of course different from the energy $A(\sim F^2)$ transferred per second from the field to the electron in the average over several collisions (equation (4)). It is this latter energy transfer which leads to a difficulty in obtaining a steady current, even in fields which are weak enough to fulfil the condition (18). This will finally lead us to the condition for breakdown.

Let us first assume a gas of free electrons which are in interaction with the lattice vibrations, and let us apply an external electrical field. Then, in the average over several collisions, the field will transfer per second the energy A (equation (4)) to each electron. In order to obtain a steady state, the electrons must be able to get rid of this energy, i.e. the energy B which an electron can transfer per second to the lattice (equation (16)) must be larger than A. Both A and B depend on the energy E of the electron (cf. figure 3).

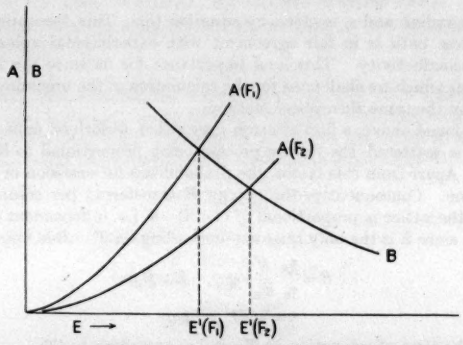


Figure 3. The energy A transferred per second from the field to the electron for two field strengths, $F_1 > F_2$, and the energy B transferred per second from an electron to the lattice as a function of the kinetic energy E of the electron.

According to equations (4), (3), (12), and (16), A increases proportionally to $E^{3/2}$, whereas B decreases as $1/E^{1/2}$. Thus for any field $F \neq 0$, there exists an energy $E'(F)$ where $A = B$,

$$A(E', F) = B(E'), \qquad \ldots.(19)$$

i.e.

$$E' \sim 1/F. \qquad \ldots.(20)$$

Electrons with an energy $E > E'$ will obtain more energy from the field than they can transfer to the lattice. Consequently, in the presence of an external electrical field, a stationary state (steady current) is impossible for electrons which interact only with the lattice vibrations. A crude picture of the movement of electrons with $E < E'$ and $E > E'$ is given in figures 4 *b* and 4 *c*.

In reality, electrons whose energy is larger than I, the lowest energy to excite or ionize an ion of the lattice, interact not only with the lattice *vibrations*, but they also make inelastic collisions with the ions. In fact, the mean free path for this latter process (in which an ion is excited or ionized) is of the order of the lattice distance, i.e. about 100 times smaller than the mean free path for scattering by lattice vibrations. This latter, therefore, can be neglected for electrons with an energy $E>I$. Consequently, the above considerations on the possibility of a steady current hold only for electronic energies $E<I$. Thus a stationary state should be impossible if $E'<I$.

In order to see in more detail what happens if $E'<I$, let us in a first approximation neglect the effect of the energy transfer A on the distribution function of the electrons. This approximation, which is generally used in calculations

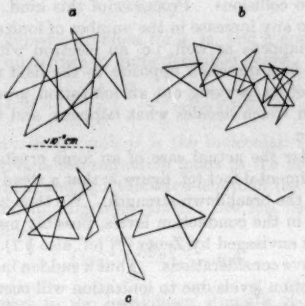


Figure 4. Crude picture of the movement of electrons: (*a*) without external field; (*b*) and (*c*) in an external field. In (*b*) the electron is slowed down, $E<E'$; in (*c*) the electron runs away, $E>E'$.

on conductivity, should be a reasonable one as long as equation (18) holds. It means that the distribution function of the electrons over the velocities is nearly independent of the directions of the velocities. In this approximation a stationary state is possible. The distribution function may, however, be different from a Maxwell distribution if $eFl \gg kT$. For $F \sim 10^6$ volt/cm., for instance, a considerable fraction of the electrons should have energies of several ev., even at room temperature.

Now let us for Δt seconds take into account the energy transfer A. Then, according to the definition of E', an electron in the energy region $E'<E<I$ will continuously gain energy from the field until it reaches the energy I and carries out an ionization. Suppose that ΔN ionization processes have been carried out after Δt seconds. Then, we have ΔN electrons less in the region $E'<E<I$, and $2\Delta N$ slow electrons more than Δt seconds ago. Thus the total number of electrons has increased by ΔN, and the original distribution function is disturbed. The first-order processes will, of course, try to restore a distribution function similar to that with which we started, i.e. the energy

region $E'<E<I$ will be refilled with electrons. But then, further ionization processes may happen through the energy transfer A, which means a continuous increase of electrons in the conduction levels. Thus a stationary state is impossible unless recombination processes are considered. Since such processes take very long compared to the times we deal with, an abrupt increase of the number of electrons in the conductive levels has to occur at the field strength at which

$$E'=I, \qquad \ldots.(21)$$

if a stationary state is possible at all.

It might be argued that even in weaker fields, when $E'>I$, it will happen that an electron having an energy nearly equal to I, and moving in the direction of the field, will be accelerated by the field in such a way that it reaches the energy I between two collisions. Processes of this kind, in the first approximation, do not lead to any increase in the number of ionization processes, since the reverse process happens as well, i.e. an electron with an energy E just above I and moving in a direction opposite to the field may lose the energy $(E-I)$ to the field before carrying out an ionization process. Thus it is the second approximation which decides what happens, and that is just what we have performed above.

Now let us consider the actual case of an ionic crystal in strong external fields. It is an experimental fact (cf. figure 2) that a steady current is obtained in strong fields below the breakdown strength. We thus have to conclude that electrons are present in the conduction levels, possibly owing to some kind of pulling-out process as envisaged by Zener [34] (cf. also § 7). To these electrons we may apply our above considerations. Thus a sudden increase of the number of electrons in conduction levels due to ionization will occur at a field strength which has to be calculated from equation (21), i.e. according to (19), from

$$A(E,F)=B(E), \quad E=I. \qquad \ldots.(22)$$

§ 5. THE BREAKDOWN FIELD[8]

From equation (22) the field strength at which a sudden increase of the number of electrons in the conduction levels occurs may be calculated. It seems to be reasonable to identify this field strength with the breakdown field F. Thus it follows from equations (22) and (4) that

$$eF=\sqrt{\left(\frac{m^*B}{\tau}\right)_{E=I}} \qquad \ldots.(23)$$

Since, according to equation (16), $B\sim 1/\tau_0$ we find

$$F\sim\frac{1}{\sqrt{(\tau\tau_0)}} \qquad \ldots.(24)$$

Thus F depends on the time of relaxation in a similar way to the electrical resistance ($\sim 1/\tau$). Actually the breakdown field depends on external parameters (as temperature, impurities) in a similar way to the electrical resistance. This will be discussed in more detail in § 6.

For ionic crystals of the NaCl type (alkali halides), we have calculated B and τ. Introducing these quantities from equations (16), (11) and (12) into (23), we obtain (with $m^*/m=1$ (cf. § 2))

$$eF=\frac{\pi^2}{2^{2/3}}\frac{(\epsilon-\epsilon_0)e^2\nu^{3/2}(mh)^{1/2}}{aI}\left[\log\frac{4\sqrt{IE_0}}{h\nu}\right]^{1/2}\phi^{1/2}(T), \quad \ldots.(25)$$

where
$$\phi(T)=1+\frac{2}{e^{h\nu/kT}-1} \quad \ldots.(26)$$

describes the temperature dependence of the breakdown field.

All quantities appearing in this expression may be measured directly. The energy I will be taken from the wave-length λ_0 of the first maximum of the ultra-violet absorption band, i.e.

$$I=hc/\lambda_0.$$

Let us introduce numerical values for e, m, and h and let us express F in v./cm. and the other quantities in units described below. We then obtain

$$F=1{\cdot}64\times10^5\left(\frac{d}{G}\right)^{1/3}(\epsilon-\epsilon_0)\frac{\lambda_0}{\lambda^{3/2}}\sqrt{\phi(T)}\ \text{v./cm.} \quad \ldots.(27)$$

Here, d is the density in gr./cm^3, G is the molecular weight, ϵ is the static dielectric constant, $\sqrt{\epsilon_0}$ is the refractive index in the near infra-red, λ is the residual wave-length in μ, and λ_0 is the wave-length of the first maximum of the ultra-violet absorption in Ångström units. Use has been made of the fact that for all alkali halides

$$\left(\log\frac{4\sqrt{IE_0}}{h\nu}\right)^{1/2}\simeq2{\cdot}6.$$

Equation (27) refers to the breakdown strength of crystals of the NaCl type only. It was possible to derive a general expression for the breakdown strength of any ionic crystal, which contains the above formula as a special case. Such a crystal has in general not only one residual frequency but several, say r, which will be denoted by $\nu_1, \nu_2, \ldots. \nu_r$. Each frequency contributes to the frequency-dependent dielectric constant $\epsilon(\nu')$ according to its oscillator strength. This contribution may be obtained from the optical constants n, κ, which according to the theory of dispersion are connected with $\epsilon(\nu')$ by

$$n^2-\kappa^2=\epsilon(\nu')=\epsilon_0+\sum_{i=1}^{r}\frac{(\epsilon_i-1)\nu_i^2}{\nu_i^2-\nu'^2}. \quad \ldots.(28)$$

Here ν' is the frequency of the light for which the optical constants are measured, and it is assumed that ν' is small compared with the first electronic absorption frequency ν_0 (in general in the ultra-violet). Consequently, if we choose ν' such that $\nu_1\ll\nu'\ll\nu_0$ we find $\epsilon_0=n^2$ in accordance with the earlier definition of ϵ_0. ϵ_i-1 is the contribution of the frequency ν_i to the static dielectric constant ϵ, which according to (28) is given by

$$\epsilon=\epsilon_0+\sum_{i=1}^{r}(\epsilon_i-1). \quad \ldots.(29)$$

A calculation shows that it is possible to obtain the mean free path of electrons in terms of ν_i and $\epsilon_i - 1$ and hence the breakdown strength F. Again, expressing F in volt/cm., $\lambda_i = c/\nu_i$ in μ and λ_0 in Ångströms, F is given by

$$F = 5{\cdot}0 \times 10^4 \left(\frac{dz}{Gr}\right)^{1/2} \lambda_0 \left\{ \sum_{i=1}^{r} \frac{\epsilon_i - 1}{\lambda_i} \phi_i(T) \times \sum_{i=1}^{r} \frac{\epsilon_i - 1}{\lambda_i^2} \log \frac{\sqrt{4E_0 I}}{h\nu_i} \right\}^{1/2} \text{v./cm.} \quad \ldots.(30)$$

Here d, G, and λ_0 have the same meanings as above, and z is the number of atoms per molecule. The temperature factor ϕ_i is obtained from equation (26) by replacing ν by ν_i. The $(\epsilon_i - 1)$ and the λ_i may be obtained from optical measurements in the infra-red, according to equation (28). Equation (30) becomes, of course, identical with equation (27) in the special case of alkali halides, where $r = 1$, $z = 2$, $(\log \gamma)^{1/2} = 2{\cdot}6$, and, cf. (29), $(\epsilon_1 - 1) = (\epsilon - \epsilon_0)$.

We thus see that a quantitative calculation of the breakdown strength of ionic crystals is possible, and we shall discuss in § 6 the formulae (27) and (30).

For non-polar crystals, such a quantitative treatment is not possible because as mentioned in § 2 the interaction between electrons and lattice vibrations is not of the simple dipolar nature that it is in polar crystals. The general ideas of the breakdown apply, however, to polar and non-polar crystals in the same way. A rough estimate of the breakdown strength may be obtained in the following way: The energy transfer B is in any crystal given by

$$B \simeq \hbar\bar{\nu} . \Phi .$$

Here $\hbar\bar{\nu}$ is the average energy transfer per collision and Φ is the probability per second that the electron makes a collision. If ϑ is the average scattering angle, the time of relaxation τ_0 is connected with Φ by

$$\Phi \simeq \frac{1}{\tau_0} \frac{1}{(1 - \cos\vartheta)} .$$

Since according to (10), for an electron with energy E

$$1 - \cos\vartheta = E/E_0,$$

where E_0 is of the order of 1 ev., we have, as in (16),

$$B \simeq \frac{\hbar\bar{\nu} E}{\tau_0 E_0} .$$

Introducing this into equation (23) and using equation (3), we find for the breakdown field

$$eF \simeq \frac{I}{l} \left(\frac{\hbar\bar{\nu}}{E_0}\right)^{1/2} . \quad \ldots.(31)$$

Since I lies usually between 5 and 10 ev., $\hbar\bar{\nu} \simeq 10^{-1}$ ev., and the mean free path l has the order of magnitude 10^{-6} cm., F must have the order of magnitude 10^6 volt/cm.

§ 6. COMPARISON BETWEEN EXPERIMENTAL AND THEORETICAL RESULTS

Breakdown strength of alkali halides at room temperature. The breakdown strength of alkali halide crystals was measured by von Hippel at room temperature. His results may be compared with the theoretical values given by equation (27). In von Hippel's experiments [18], the electrical field lies in the 100 direction of the crystals, whereas breakdown usually occurs in the 110 direction. This suggests that $F_{110} : F_{100} = 1 : \sqrt{2}$, where F_{110} and F_{100} are the respective breakdown strengths when the directions of the field are the 110 and the 100 direction respectively. There is no direct experimental evidence for this, but von Hippel gives an experimental example for a similar case (i.e. for another direction of the breakdown). Accordingly we shall multiply the F-values of equation (27) by a factor $\sqrt{2}$.

This dependence on the direction means that we have to assume $m_{110} < m_{100}$ and $m_{110} \sim m$, where m_{110} and m_{100} are the effective masses in the 110 and 100 directions, respectively. For very small energies of an electron, this would not be possible because of the cubic symmetry of the crystal. The energies near I (5–10 ev.), however, lie in most cases in a higher Brillouin zone, where such a behaviour might very well be expected, since high effective mass means a potential barrier between the atoms, and since, as pointed out by von Hippel, the potential barrier in the 100 direction is much higher than that in the 110 direction.

In table 1 we compare the experimental and theoretical values of the breakdown strength at room temperature. The experimental data for ϵ, ϵ_0 and λ may be taken from a table given in the article by Born and Göppert-Mayer [3] (for ϵ the values of Erréra have been used) and λ_0 from the measurements by Hilsch and Pohl [15].

Table 1. F_{100} *in* 10^5 v./*cm. for* $T = 300°$ K.

	NaCl	NaBr	NaI	KCl	KBr	KI	RbCl	RbBr	RbI
Experiment	15	10	8	8	7	6	7	6	5
Theory	15	13	14	8·9	7·3	6·3	8·4	5·4	4·7
$\epsilon - \epsilon_0$	3·4	3·3	3·6	2·6	2·4	2·3	3·0	2·4	2·2
λ in μ	61	74	86	71	88	102	85	115	129
λ_0 in A.	1580	1900	2295	1625	1875	2190	1670	1920	2225

The agreement must be considered as satisfactory, in view of the fact that no arbitrary constants enter into the theoretical formula.

Quartz and mica. In contrast to the alkali halides, both quartz and mica have more than one residual wave-length. Consequently, the breakdown strength has to be calculated from the general formula (30). Quartz has four residual wave-lengths. The two strongest are $\lambda_1 = 8{\cdot}4$ and $\lambda_2 = 20{\cdot}5$.

Unfortunately, the available optical data are not sufficient to deduce the relative strengths (ϵ_1-1) and (ϵ_2-1) of the oscillations. It seems, however, that the assumption that the oscillator strengths of λ_1 and λ_2 are equal $(\epsilon_1=\epsilon_2)$ and that the other two residual wave-lengths may be neglected is not in contradiction to experiment. This leads (with $\epsilon=4{\cdot}6$, $\epsilon_0=2{\cdot}3$, $\lambda_0=1200$ A.) to $F=3{\cdot}3\times 10^6$ v./cm. On the other hand, the assumption that only the shortest residual wave $\lambda_1=8{\cdot}4\mu$ is of importance gives an upper limit for F, i.e. $F<7{\cdot}4\times 10^6$ v/cm. The experimental value, kindly communicated by the E. R. A., is $6{\cdot}7\times 10^6$ v./cm. for crystalline quartz.

For mica the strong residual wave-lengths are about the same as for quartz, but even less information than for quartz is available about their relative strengths. As in the case of quartz, however, it is possible to give an upper theoretical limit, which is 15×10^6 v./cm. Experimental values, communicated by the E. R. A., lie between 10 and 11×10^6 v./cm. and up to 14×10^6 v./cm. with impulse voltages as communicated by the National Physical Laboratory.

Dependence on temperature. According to theory, the intrinsic electrical strength of ionic crystals should increase with increasing temperature * (cf. equations (26), (27) and (30)). This is due to an increased scattering of electrons with increasing temperature. According to equation (26) the temperature dependence should be the stronger, the larger the residual wave-length. In table 2 we give for different ν the value of the quantity $\sqrt{\phi(T)}$ for $T=300°$, i.e. the ratio of the theoretical breakdown strength at room temperature and at the absolute zero of temperature. Thus most alkali halides with their large residual wave-lengths should show a considerable dependence on temperature. Quartz and mica, however, with their small residual wave-lengths should have a breakdown strength nearly independent of temperature (at least up to the temperature T_0, cf. first footnote below).

Table 2. $\sqrt{\phi(T)}$ *at* $T=300°$ K.

λ in μ	10	20	30	40	50	60	80	100	120
$\sqrt{\phi(300)}$	1·01	1·10	1·23	1·37	1·51	1·64	1·87	2·08	2·28

Before this theory was put forward, it was generally assumed that the breakdown strength does not increase with temperature but rather decreases, as was found for more complicated substances. In order to check the theoretical results, experiments on mica and KBr were carried out by Austen and Hackett [1] at the E. R. A. laboratory. The results are shown in figures 5 and 6. We see that for mica the electric strength is nearly independent of the temperature, as it should be according to table 2, since its strong residual wave-lengths are $8{\cdot}4\ \mu$ and $20{\cdot}5\ \mu$. For KBr ($\lambda=88\ \mu$) a strong decrease of the electric strength with decreasing temperature was found, in qualitative but not quantitative agreement with theory. Von Hippel [18] independently obtained a qualitatively similar result.†

* This holds only up to a temperature T_0 (of the order of 100° C.) where a complete change in the breakdown mechanism begins (heat breakdown, cf. Part II).

† Kindly communicated to the author before publication.

Since the experiments by Austen and Hackett and by von Hippel are not in quantitative agreement with each other, it seems to be premature to draw too far-reaching conclusions. There is a qualitative agreement, however, with respect to the fact that the electric strength decreases with decreasing

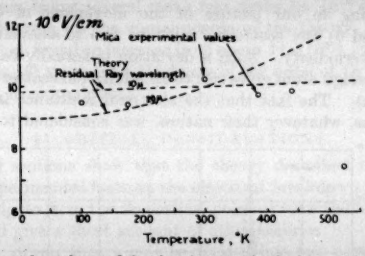


Figure 5. Temperature-dependence of the electric strength of mica. o, experiments by Austen and Hackett [1]. - - -, theory for $\lambda = 10\,\mu$ and $\lambda = 20\,\mu$.

temperature,* and that this effect is much more pronounced for KBr with its large λ than for mica which has much smaller residual wave-lengths.

Mixed crystals. In § 5 we mentioned that the electric strength depends on the mean free path in a similar way to the electrical resistivity of metals.

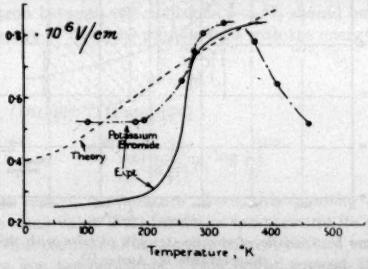


Figure 6. Temperature-dependence of the electric strength of KBr. - - - - -, theory. ———, measurements by Austen and Hackett [1]. —o— - —, measurements by von Hippel [18].

Thus, for instance, both increase with temperature, owing to an increased scattering of the electrons.

A similar analogy is observed in the presence of foreign atoms. It is well known that the electrical resistance of metals is always increased by foreign

* At high temperatures, in mica as well as in KBr, a decrease of the electric strength with increasing temperature is found. This should be due to heat breakdown, cf. Part II and footnote on p. 424.

atoms (Matthiessen's rule). On the other hand, von Hippel observed that the electric strength too is always increased by foreign atoms (cf. figure 1).

From the theoretical point of view,* the increase of the resistance or of the electric strength must be due to an additional scattering by the foreign atoms. According to our picture of the movement of electrons through crystals (§§ 2 and 3) any scattering must be due to a deviation of the lattice potential from periodicity. Such a deviation is actually caused by a foreign atom, i.e. any foreign atom gives rise to an additional scattering, even if at rest in a lattice point. The fact that the electrical resistance is always *increased* by foreign atoms, whatever their nature, was considered to be an important

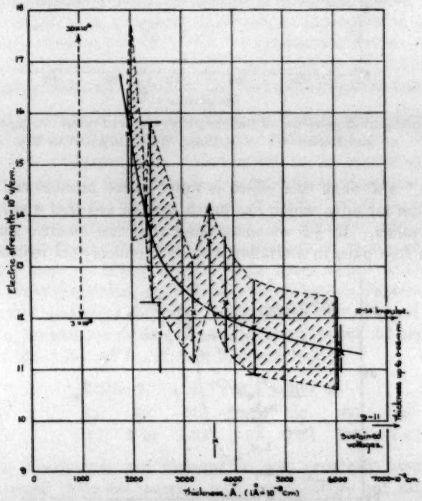


Figure 7. Variation of electric strength of mica with thickness. Experiments by Austen [1].

confirmation of the basic ideas of the scattering of electrons in metals. Since we have used the same ideas, von Hippel's experimental results may be considered as an important confirmation of our ideas on electrical breakdown.

In view of the importance of the effect, it was desirable to have quantitative theoretical results for the increase of the electric strength. These could be obtained for the dilute solid solution [10] of RbCl in KCl, using a method which was developed by Mott [19] for the case of metals. The relative increase of the electric strength was found to agree within 30 per cent with the experimental value, an agreement which must be considered as very satisfactory.

* The theoretical derivation of Matthiessen's rule is due to Nordheim [21].

Dependence on thickness. The breakdown strength of thin dielectric layers should, according to theory, increase with decreasing thickness. A quantitative theoretical treatment of this effect would be rather difficult. Qualitatively it can be said that an appreciable increase of the electric strength should be found at thicknesses below several mean free paths, i.e. 10^{-5} cm. Austen and Hackett [1], in following up the consequences of the theory, have found this effect in mica. They found an increase from about 11×10^6 v./cm. at a thickness of 10^{-4} cm. to 16×10^6 v./cm. at a thickness of the order of 10^{-5} cm. Figure 7 shows their result in more detail.

§ 7. CRITICAL CONSIDERATIONS

The preceding sections show that the theory described here gives a fair account of the experimental facts on the electrical breakdown. In this section we want to investigate from the theoretical point of view the basic assumptions of the theory, and give a short account of other theories.

The discussion of our theory may be divided into two parts : (i) the calculation of the mean free path l and of the energy transfer B to the lattice (§ 3), and (ii) the condition for breakdown (§ 4). The calculation of l and B was carried out with the help of well-known methods, and there seems to be little doubt that for the electronic energies in question (5–10 ev.) they lead to a reasonable first approximation. The condition for breakdown, as discussed in § 4, is a reasonable condition as long as equation (18) is fulfilled for the breakdown field F. This means that even at the breakdown field, the energy change of an electron between two collisions ($\sim eFl$) should be small compared with its kinetic energy. Thus, for the electrons with the energy $E=I$, we must have

$$eFl/I \ll 1,$$

where, using (23), (16), (3), (11) and (26),

$$\frac{eFl}{I} = \left[\frac{2\hbar\nu}{E_0\phi(T)} \log \gamma\right]^{1/2} \qquad \ldots.(32)$$

At the absolute zero of temperature, $T=0$, this quantity has for all alkali halides nearly the same value 0·4. At room temperature its value is between 0·2 and 0·3. We thus see that eFl is smaller than I, but not very much smaller, especially at very low temperatures. This might suggest that such details of the theory as the dependence on temperature are correct only in a qualitative sense.

Von Hippel [16] and Seeger and Teller [24], [25] have proposed another condition for the calculation of the breakdown field. They propose that the field must be so strong that electrons with an *arbitrary* velocity gain from the field more energy than they transfer to the lattice. Now this energy transfer is a quantity which is zero for electronic energy $E=0$, then rises to a maximum for electrons with an energy of several $\hbar\nu$ and for higher energies falls continuously. Thus this condition would give a higher breakdown strength than ours. How much higher it would be is difficult to say at present, since an accurate calculation

of the energy transfer for electronic energies of several $h\nu$ seems to be very difficult.*

A mechanism of a quite different kind from those considered till now was proposed by Zener [34]. He finds that in strong fields there is a chance that an electron of a lower energy band may penetrate a forbidden energy region and thus come into the conduction band. The probability of such a penetration increases rapidly with the field strengths. Processes of this kind are possibly of importance for the setting up of an electronic current before breakdown. They could not, however, account for the sudden instability, as already noted by Whitehead [31]. Other effects too, such as the influence of temperature and of foreign atoms, could not be explained by such a theory.

II. BREAKDOWN THROUGH THERMAL INSTABILITY †

The intrinsic electric strength determines, as discussed in § 1, the highest electrical field which can be applied to a dielectric. Even in weaker fields, however, thermal instabilities may occur under certain circumstances, which ultimately lead to a breakdown. Thermal instability is to a large extent due to the ionic conductivity of crystals and, therefore, is usually of importance only at high temperatures.

The theory of breakdown through thermal instability was first given by Miles Walker [29] in 1912. Expressed analytically his treatment is described by the fundamental equations

$$W = H, \qquad \frac{\partial W}{\partial \theta} = \frac{\partial H}{\partial \theta}, \qquad \ldots.(33)$$

where W = energy generated in the dielectric by reason of its dielectric losses or conductivity.

H = rate of dissipation of heat from the dielectric.

θ = temperature of dielectric.

Now if $\partial^2 W/\partial\theta^2$ exceeds $\partial^2 H/\partial\theta^2$, the preceding equations give the condition of tangency between the curves of generation (W) and dissipation (H) of heat, the former being more convex to the temperature axis than the latter. Accordingly, if the equilibrium is disturbed by an increase in W or θ the temperature of the dielectric will increase indefinitely, i.e. it is thermally unstable. Since W is a function of the applied voltage, the solution of the equations gives a " critical " or " minimum breakdown " voltage, which is, of course, of practica limportance only if lower than the intrinsic strength. The first detailed treatment of a special case was given by Wagner [28].

Equations (33) do not take into account the temperature distribution in the dielectric. Fock [6] formulated the general equations for the limit of

* It is doubtful whether this has already been done in a satisfactory way in the first paper by Seeger and Teller [24] as discussed in later papers [9], [25]. Also it seems difficult to see how many of the observed facts (temperature dependence, effect of impurities) could be explained.

† This Part is mainly due to Dr. S. Whitehead, whom I also wish to thank for many discussions. For a detailed report up to 1932, cf., for example, the book by S. Whitehead [30].

thermal equilibrium and applied these to some special cases ; further special cases have been given by Dreyfus [5], Rogowski [22], Brazier [4], Whitehead [32], etc. One deduction from Fock's analysis is that, subject to the conductivity being independent of the field strength and to certain other conditions, the electric strength should vary with temperature in the same manner as the square root of the electrical resistivity. This has been verified experimentally for a number of dielectrics, notably by Inge and Walther, who showed that this law was obeyed unless the breakdown voltage corresponding to thermal instability was so high that other types of breakdown (intrinsic strength) supervened. The work of Inge and Walther was very extensive, so that reference is made to the book by Semenoff and Walther [26], where a summary will be found.

Another deduction is that, considering a plane slab of dielectric of infinite area, the breakdown voltage tends to an asymptotic value as the thickness is indefinitely increased, or, under certain conditions, if the surface is maintained at the ambient temperature. This law in its first form is a very general one and requires merely that the energy losses in the dielectric should increase with temperature and that there should exist a limiting temperature beyond which the dielectric loses its insulating properties. This limiting voltage is a function only of the properties of the dielectric and represents the greatest voltage which can be supported by a slab of the dielectric as described. The value is given as the solution of the equation :

$$V^2 = \int_{\theta_0}^{\theta_s} \frac{8\lambda}{\sigma}\, d\theta, \qquad \ldots.(34)$$

where λ=thermal conductivity, σ=electrical conductivity, V=limiting voltage, θ=temperature, θ_s=limiting temperature, θ_0=ambient temperature.

The values of this voltage are very high (10^6–10^{10} volts for D.C. and A.C. of industrial frequency), but may be low for commercial dielectrics at radio frequency, cf. [31].

Dreyfus [5], Whitehead and Nethercot [33], and Gemant and Whitehead [12] have extended the analysis to those cases where the conductivity is a function of the field strength. The two latter papers show that, in most cases, if E_m is the electric strength when the electrodes are thermally insulated in such a way that there is a negligible temperature gradient in the specimen, and if E is the electric strength for the actual electrode arrangement, then E/E_m is a universal function of the ratio of the thermal resistance of the electrodes to the thermal resistance of the specimen. This function may easily be deduced from Fock's analysis but applies much more generally. E_m is calculated from (33), which is easy if the functions are explicit.

Thermal instability precedes breakdown by an interval which decreases as the voltage is raised beyond the minimum value necessary to produce instability. This has been considered by Wagner [28] and Whitehead and Nethercot [33], the latter having verified experimentally the deduction that the effect of variables such as temperature is unchanged provided a fixed or a sufficiently large time interval is adopted. More recently Hartree [13] has

shown how more general solutions of this problem may be obtained by the differential analyser.

To conclude, thermal instability has been reduced to a mathematical computation, which is theoretically always possible. The range of application is, however, on account of the interposition of other mechanisms, limited to especially favourable circumstances such as high ambient temperature, low thermal dissipation, high frequency, great thickness and poor insulating properties.

REFERENCES

(1) Austen and Hackett. *Nature, Lond.*, **143**, 637 (1939); Austen and Whitehead (unpublished).
(2) Bloch. *Z. Phys.* **52**, 555 (1928).
(3) Born and Göppert-Mayer. *Handb. d. Physik*, 2nd ed. **24** (2), 623 (1933).
(4) Brazier, *J. Instn. Elect. Engrs*, **27**, 104 (1935).
(5) Dreyfus. *S.E.V. Bulletin*, nos. 7 & 15 (1924).
(6) Fock. *Arch. Elektrotech.* **19**, 71 (1927).
(7) Fröhlich. *Elektronentheorie der Metalle* (Berlin, 1936).
(8) Fröhlich. *Proc. Roy. Soc.* A, **160**, 230 (1937); **172**, 94 (1939).
(9) Fröhlich. *Phys. Rev.* **56**, 349 (1939).
(10) Fröhlich. *Report to the E.R.A.* (in the press).
(11) Fröhlich and Mott. *Proc. Roy. Soc.* A, **171**, 496 (1939).
(12) Gemant and Whitehead. *Phil. Mag.* **27**, 582 (1939).
(13) Hartree. *J. Instn. Elect. Engrs*, **85**, 56 (1939).
(14) Hartshorn. *Rep. Progr. Phys.* **4**, 218 (1937).
(15) Hilsch and Pohl. *Z. Phys*, **59**, 812 (1930).
(16) von Hippel. *Ergebn. exakt. Naturw.* **14**, 79 (1935). With references to von Hippel's earlier papers.
(17) von Hippel. *Phys. Rev.* **54**, 1096 (1939).
(18) von Hippel. *Phys. Rev.* **56**, 941 (1939).
(19) Mott. *Proc. Camb. Phil. Soc.* **32**, 281, 19 (1936).
(20) Mott and Jones. *The Structure of Metals and Alloys* (1936).
(21) Nordheim. *Ann. Phys., Lpz.*, **9**, 607 (1931).
(22) Rogowski. *Arch. Elektrotech.* **13**, 155 (1924).
(23) Rogowski. *Sommerfeld-Festschrift* (*Probleme der modernen Physik*). (Leipzig, 1928.)
(24) Seeger and Teller. *Phys. Rev.* **54**, 515 (1939).
(25) Seeger and Teller. *Phys. Rev.* **56**, 352 (1939).
(26) Semenoff and Walther. *Die physikalischen Grundlagen der elektrischen Festigkeitslehre.* (Berlin, 1928.)
(27) Tibbs. *Trans. Faraday Soc.* (December 1939).
(28) K. W. Wagner. *J. Amer. Inst. Elect. Engrs*, **12**, 1034 (1922).
(29) Walker, M. *Theory of Machine Design* (1912).
(30) Whitehead, S. *Dielectric Phenomena*, **3**, 299. (London, 1932.)
(31) Whitehead, S. *Rep. Brit. Ass.* (1936).
(32) Whitehead, S. *Phil. Mag.* **27**, 276 (1939).
(33) Whitehead, S. and Nethercot. *Proc. Phys. Soc.* **47**, 974 (1935).
(34) Zener. *Proc. Roy. Soc.* A, **145**, 523 (1934).

THE TEACHING OF PHYSICS IN TECHNICAL INSTITUTIONS

By D. OWEN, B.A., D.Sc., F.Inst.P.

The term Technical Institution covers a great variety of schools and colleges in Great Britain, from the junior technical school to the college of university rank, such as the Imperial College of Science and Technology. In all alike the emphasis is laid on the study of those branches of knowledge which are of special value in relation to industry. With regard to scientific studies, it may be generally laid down that the scope of the subject-matter and the methods of teaching are almost indistinguishable whether, at the elementary stage, they are given in a technical school or in a secondary school; or whether, in the more advanced stages, they are given in a college of a university or in a higher technical institution.

Scientific discovery and invention led to that great expansion in the scope and methods of industry in this country which began in the eighteenth century and which, in the language of some writers, has been designated the Industrial Revolution. With this background the first Mechanics' Institutes were founded by George Birkbeck over a century ago, the object in view being the dissemination of a knowledge of scientific principles amongst the workers and artisans. But it was soon discovered that there is no short cut to an effective specialized knowledge of the sciences in application to any particular branch of industry. It became apparent that a wider educational basis was necessary, including a certain command of language and a fundamental and systematic training in the relevant branches of knowledge. In the last decade of the nineteenth century the rapidly widening realization of the importance of a scientific training for industry resulted in the establishment of large and well-equipped technical institutions throughout the country. The change in the point of view is illustrated by the cases of the Mechanics' Institute founded in London by Birkbeck in 1823, which in 1920 became a school of the University of London, and that founded by him in Glasgow earlier in the same year, which has since become the Royal Technical College and part of Glasgow University.

We are here concerned with one subject only in the scientific curriculum, namely, physics. In the junior technical school this subject is carried to a stage which may be sufficiently described as that of the school-leaving certificate. In the "Evening Institutes" (covering the age range 14–16) the work undertaken is of a quite elementary character, in which physics appears only incidentally and superficially. In a large number of technical institutes which may be placed in an intermediate category, though a definite and systematic scientific training involving lectures and laboratory work is given, physics plays only a subsidiary part; their students are usually satisfied

to work for diplomas in the form of National Certificates, which are awarded in chemistry and in engineering, but not in physics. The syllabuses and examinations for such certificates are arranged jointly by the Institutes and the Board of Education.

The thoroughgoing study of physics is confined to the large technical institutes in London and in the provinces. These higher technical institutions, while under the immediate direction of a governing body, are largely controlled by the Board of Education and the local education authorities. Systematic courses in physics are provided in these institutes, from the elementary up to advanced stages comparable to those given in the colleges of a university. The remainder of this article deals with the teaching of physics in technical institutions falling in this category.

The students in the higher technical institutions may be divided into those taking full-time day courses, who form only a small minority; and those taking part-time courses, which are, in some institutes, held during the morning or afternoon as well as in the evening, though the great majority of students are able to attend during the evenings only. Part-time students attending during the day periods consist to a large extent of apprentices who are set free under special arrangements made by their firms and the authorities of the school: generally speaking the ground covered in physics does not extend much beyond the elementary stage, and the course may extend over one or two years.

The elementary or first-stage courses are, as a general rule, taken by students who come from an elementary school, or whose interest in science has been aroused only after leaving school. The handicap of an insufficient grounding in the subjects of a general education during the school years is difficult to shake off, and progress in after-study is often slow. Only a minority are able to pass the matriculation examination, which is the usual prelude to the serious pursuit of any special subject. Notable exceptions occur, when by inherent ability and character the student triumphs over these initial obstacles and is able to proceed effectively to higher studies and even to research. Those less fortunately endowed have usually to be content to come within the category of those for whom physics is merely ancillary to a course in some other main subject, in preparation, say, for a National Certificate.

The boy or girl who has passed through a secondary school is in a much more favourable position to take up a scientific course of study. He has normally obtained his school-leaving certificate and perhaps qualified for matriculation; or he may have continued his school course in preparation for the higher certificate examination and obtained exemption from the university intermediate examination. Unless he is one of the minority who are able to proceed to a university he may enter on a career in industry. The continuation of his studies in any selected direction is now normally restricted to the evening. That the urge to such continuation of studies exists is shown by the great demand for courses of the same general character as those that would have been open to them had they been fortunate enough to pass direct from school to university. It has also to be recognized that

the choice of going up from school to university is often decided in the negative not by financial inability, but in deference to the view, held by many employers, that entry into the professions or into the higher ranks of industry should not be delayed beyond the age of about eighteen. For various reasons there thus exists a large category of students well qualified and eager on leaving school to pursue their scientific studies to an advanced stage during their apprenticeship within a profession or industry. A great and steady demand has accordingly been made on the higher technical institutions for courses of education in the various branches of science and mathematics up to the standard of a university degree. Adequate facilities in staff and equipment to meet this demand are available at these institutions, which thus serve a function parallel to that of a university.

The work carried on in technical institutions has often been designated by such terms as "technical instruction", or "vocational training", but, in fact, as already indicated, such instruction or training, to be adequate to the purpose, must be so broadly conceived that it is in the main little distinguishable from what is usually termed academic education. Reference is naturally constantly made by his teacher to examples of application in an instrument or a machine or a process; but it is generally agreed that lengthy digressions from the broad course of exposition of physical principles have to be guarded against. Such diversions may well prove a source of loss rather than of gain to the student's command of his subject and his creative ability to use his knowledge in the intensive range of application which belong to his own particular present or prospective line of work in the industrial world.

The student naturally aims at securing some recognized diploma as a guarantee of his qualifications and a seal to his application and ability. His nominal objective is generally to obtain a university degree, which may be of a pass or an honours standard, and may be followed later, through research, by a higher degree, either M.Sc. or Ph.D. or D.Sc. The Institute of Physics, a corporate body of professional physicists, recognizes the attainment of a good university degree as the normal preliminary condition for award of a diploma of corporate membership. The entrants on degree courses of study at a higher technical institution will usually have already had a good preliminary secondary school training in physics, often up to the standard of a university intermediate examination. The University of London is notable as having provided a road to a university degree available to students who are off the normal track of university education. In the provinces the external degree is the objective, but it is possible for the student living in or near London to take an internal degree or, in other words, in respect of examinations, to enjoy to an appreciable extent the privileges open to members of a School of the university. This status of association with London University is one which is equally prized by the teachers "recognized" for the purpose by the University and by the registered internal students who attend courses fulfilling conditions laid down by the University. Both on the internal side and on the external side the provisions which this University has made for the benefit of students occupied in earning their living have met with a success which is widely recognized.

Some particulars may be given in regard to the duration of courses of work undertaken by part-time students. Newly matriculated students usually qualify for a pass in the Intermediate degree examination in science at the end of two years. The course for the "pass" B.Sc. degree requires a further two or, more usually, three years' study; whilst for a "special honours" degree the required period after the intermediate stage extends to three or preferably four years. The classes include a sprinkling of students who have already taken a normal university course and obtained a degree, and who desire, under the limitations of part-time study, to proceed to a higher degree or to specialize in a new branch of science.

Post-graduate study at a higher technical institution may take one of two forms. It may consist of attendance at special short courses of about ten lectures, sometimes accompanied by laboratory work. The courses may be given by members of the full-time staff of the institute, or by specialists associated with scientific firms or with research associations. As examples of the subject-matter of such special courses may be mentioned architectural acoustics, colloid physics and electric-wave propagation. Post-graduate study may also take the form of research, obviously a somewhat formidable undertaking for a part-time student. The demand for facilities for research at these institutions is steady and considerable. It may come from students who have distinguished themselves in their first degree examinations after a regular course at one of these institutions, or from those who have graduated elsewhere, often after a normal university course. Success in research conducted towards the close of an already strenuous day is certainly difficult, but it is not impossible. The higher technical institutions can point with pride to a long list of published researches which have not merely culminated in the acquisition of a higher degree, but have resulted in valuable additions to the common stock of scientific knowledge.

Mention may be allowed of less formal activities, such as those of student physical societies, in technical institutions. Apart from social intercourse, such societies afford opportunities to students to prepare and read papers of physical interest at the society's meetings, serving to excite their interest in the expansion of scientific frontiers and, incidentally, to mark out students of outstanding ability. Visits are arranged to works, research laboratories and the like. Addresses are occasionally given by distinguished visitors: these meetings are often very memorable occasions which leave a lasting inspiration, and no gratitude can be too warmly expressed for those men of distinction who, from time to time, give up an evening, and give of their best, in response to invitations from these societies.